ARL-TR-7925 • JAN 2017

US Army Research Laboratory

Quantal Response: Nonparametric Modeling

by Joseph C Collins

NOTICES

Disclaimers

ARL-TR-7925 • JAN 2017

US Army Research Laboratory

Quantal Response: Nonparametric Modeling

by Joseph C Collins
Survivability/Lethality Analysis Directorate, ARL

REPORT DOCUMENTATION PAGE			*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)* January 2017	**2. REPORT TYPE** Technical Report	**3. DATES COVERED (From - To)** October 2013–December 2016
4. TITLE AND SUBTITLE Quantal Response: Nonparametric Modeling		**5a. CONTRACT NUMBER**
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Joseph C Collins		**5d. PROJECT NUMBER**
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Army Research Laboratory ATTN: RDRL-SLB-D Aberdeen Proving Ground, MD 21005-5068		**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TR-7925
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		**10. SPONSOR/MONITOR'S ACRONYM(S)**
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Author's Email: <joseph.c.collins38.civ@mail.mil>

14. ABSTRACT
Binary response to a continuous stimulus was originally modeled as a cumulative distribution of limit quantities interpreted as an increasing functional relationship between stimulus and probability of response. The Generalized Linear Model approach does not make use of the limit distribution but allows arbitrary functional relationships including both classical higher-order polynomials and also nonparametric spline models based on penalized maximum likelihood estimation.

15. SUBJECT TERMS
quantal response, polynomial, penalized likelihood, smoothing, spline

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Joseph C Collins
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UU	70	**19b. TELEPHONE NUMBER (Include area code)** 410-278-6832

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures iv

List of Tables v

1. Introduction 1

2. General QR Models 1

3. The Chang-Bodt QR Model 2

4. Parametric QR Models 5

5. Nonparametric QR Models 8

5.1 The P-spline QR Model 8

5.2 The S-spline QR Model 12

6. Model Comparison 16

7. Conclusions and Recommendations 18

8. References 19

Appendix A. The Linear Model 21

Appendix B. The Generalized Linear Model 33

Appendix C. B-splines 47

Appendix D. Matrix Differentiation Operators 53

List of Symbols, Abbreviations, and Acronyms 58

Distribution List 59

List of Figures

Fig. 1 Chang-Bodt model 4
Fig. 2 Regression basis sets 6
Fig. 3 Logistic regression 7
Fig. 4 P-spline optimization 9
Fig. 5 Optimal and adjusted P-spline solutions 10
Fig. 6 Adjusted P-spline solution with 90% confidence intervals 11
Fig. 7 S-spline optimization 13
Fig. 8 Optimal S-spline solutions 14
Fig. 9 Adjusted S-spline solution with 90% confidence intervals 15
Fig. 10 All solutions 17
Fig. A-1 P-spline optimization, linear model 26
Fig. A-2 S-spline optimization, linear model 28
Fig. A-3 Smoothing parameter selection, linear model 30
Fig. A-4 GCV-optimal fits, linear model 31
Fig. B-1 P-spline optimization, GLM 40
Fig. B-2 S-spline optimization, GLM 41
Fig. B-3 Smoothing parameter selection, GLM 42
Fig. B-4 GCV-optimal fits, GLM 43
Fig. C-1 B-spline basis, uniform 49
Fig. C-2 B-spline basis, compact 50
Fig. C-3 B-spline basis derivatives, uniform 51
Fig. C-4 B-spline basis derivatives, compact 52

List of Tables

Table 1 Chang-Bodt model parameters 3

Table 2 Goodness of fit comparison 16

INTENTIONALLY LEFT BLANK.

1. Introduction

The quantal response (QR) model seeks to characterize the value of a binary response $y \in \{0,1\}$ as a function of continuous stimulus x. The original formulation postulates an unobservable stimulus limit L, which determines the response via

$$y = \begin{cases} 0, & x < L \\ 1, & x \geqslant L\,. \end{cases} \tag{1}$$

If L is known and constant, then y is a step function of x with jump at $x = L$. Otherwise, suppose L is random with cumulative distribution function (CDF)

$$F_L(t) = \Pr[\, L \leqslant t\,]\,. \tag{2}$$

The distribution of L determines the QR model probability of response as

$$P(x) = \Pr[\, y = 1 \mid x\,] = \Pr[\, L \leqslant x\,] = F_L(x)\,. \tag{3}$$

Assuming a specific distribution family for F_L such as logistic, $G(x) = (1+e^{-x})^{-1}$, or normal, $G(x) = (2\pi)^{-1/2}\int_{-\infty}^{x} e^{-t^2/2}\,dt$, fitting the model amounts to estimating the limit distribution location and scale parameters m and s in

$$P(x) = F_L(x) = G\left(\frac{x-m}{s}\right)\,. \tag{4}$$

The CDF of L is interpreted as a functional model for the probability of response.

The existence of L is not necessary, and as such the QR model is a special case of the Generalized Linear Model (GLM). See Collins[1,2] for details. The requirement of increasing P can be relaxed, and then P is no longer equivalent to any CDF (which, of course, must always be increasing). What remains is a purely functional model with no implicit limit distribution. Several such formulations follow.

2. General QR Models

A binary response $y \in \{0,1\}$ has an expected value depending on stimulus x,

$$\mathrm{E}[\, y \mid x] = P(x)\,. \tag{5}$$

The conditional distribution of y is Bernoulli with the indicated mean, so

$$\Pr[y = 1 \mid x] = E[y \mid x] = 1 - \Pr[y = 0 \mid x] . \tag{6}$$

In the usual case, response is an increasing function of x, so the response function P must be a CDF. A standard choice for P uses the logistic CDF $G(x) = (1 + e^{-x})^{-1}$ with the linear parameterization $P(x) = G(a + bx)$ or the location-scale parameterization $P(x) = G((x - m)/s)$.

This is inadequate, for example, in the presence of the shattergap phenomenon, when the probability of penetration (y) is observed to decrease in some velocity (x) range. An example data set is taken from Chang and Bodt,[3] described therein as "Results of 69 Ballistic Shots on Phase II Al_2O_3/Kevlar Armor Plates", although there are only 68 data points in the report.

Chang and Bodt also develop a specific parametric model that must be analyzed from first principles (not GLM). This is presented in Section 3.

The remaining approaches are all based on GLM. Standard Bernoulli GLM can estimate an arbitrarily complex response, not necessarily increasing, as in Section 4. Nonparametric estimation of the response is accomplished with penalized B-spline models, as in Section 5.1, or smoothing spline models, as in Section 5.2.

3. The Chang-Bodt QR Model

The Chang-Bodt[3] model is

$$P(x) = (1 - P_z(x, m_z, s_z)) \cdot P_1(x, m_1, s_1) + P_z(x, m_z, s_z) \cdot P_2(x, m_2, s_2) , \tag{7}$$

where the P_i are location-scale CDFs. P_1 is the (monotonic) probability of penetration for an unshattered threat, and P_2 is the probability of penetration for a shattered threat. P_z is the probability of shatter. For low x, shatter is unlikely, $P_z \sim 0$, and $P \sim P_1$. For high x, shatter is likely, $P_z \sim 1$, and $P \sim P_2$. In the intermediate x range, the mixture is weighted according to the increasing shatter probability P_z. For convenience, let $R = 1 - P$.

The likelihood for a single observation (x_i, y_i) is

$$L_i = R_z(x_i) \cdot P_1(x_i)^{y_i} R_1(x_i)^{1-y_i} + P_z(x_i) \cdot P_2(x_i)^{y_i} R_2(x_i)^{1-y_i} , \tag{8}$$

where the parameters of P_k are (m_k, s_k).

$$L_i = \begin{cases} R_z(x_i) \cdot R_1(x_i) + P_z(x_i) \cdot R_2(x_i) , & y_i = 0 \\ R_z(x_i) \cdot P_1(x_i) + P_z(x_i) \cdot P_2(x_i) , & y_i = 1 \end{cases} . \tag{9}$$

For the logistic CDF, we have $G(x) = (1 + \exp(-x))^{-1}$. The usual location-scale parameterization is $P(x) = G((x - m)/s)$.

Parameter estimates can be obtained by numerical optimization of the negative log-likelihood $\Lambda = -\sum \log L_i$. The parameters reported by Chang and Bodt are not optimal estimates. Both sets of parameter estimates are given in Table 1.

Table 1 Chang-Bodt model parameters

Model	m_1	s_1	m_2	s_2	m_z	s_z	Λ
Reported	1650	100	2550	100	2050	100	22.58
Estimated	1646.97	90.49	2529.34	75.75	2020.15	87.80	21.91

See Fig. 1 for the response curves.

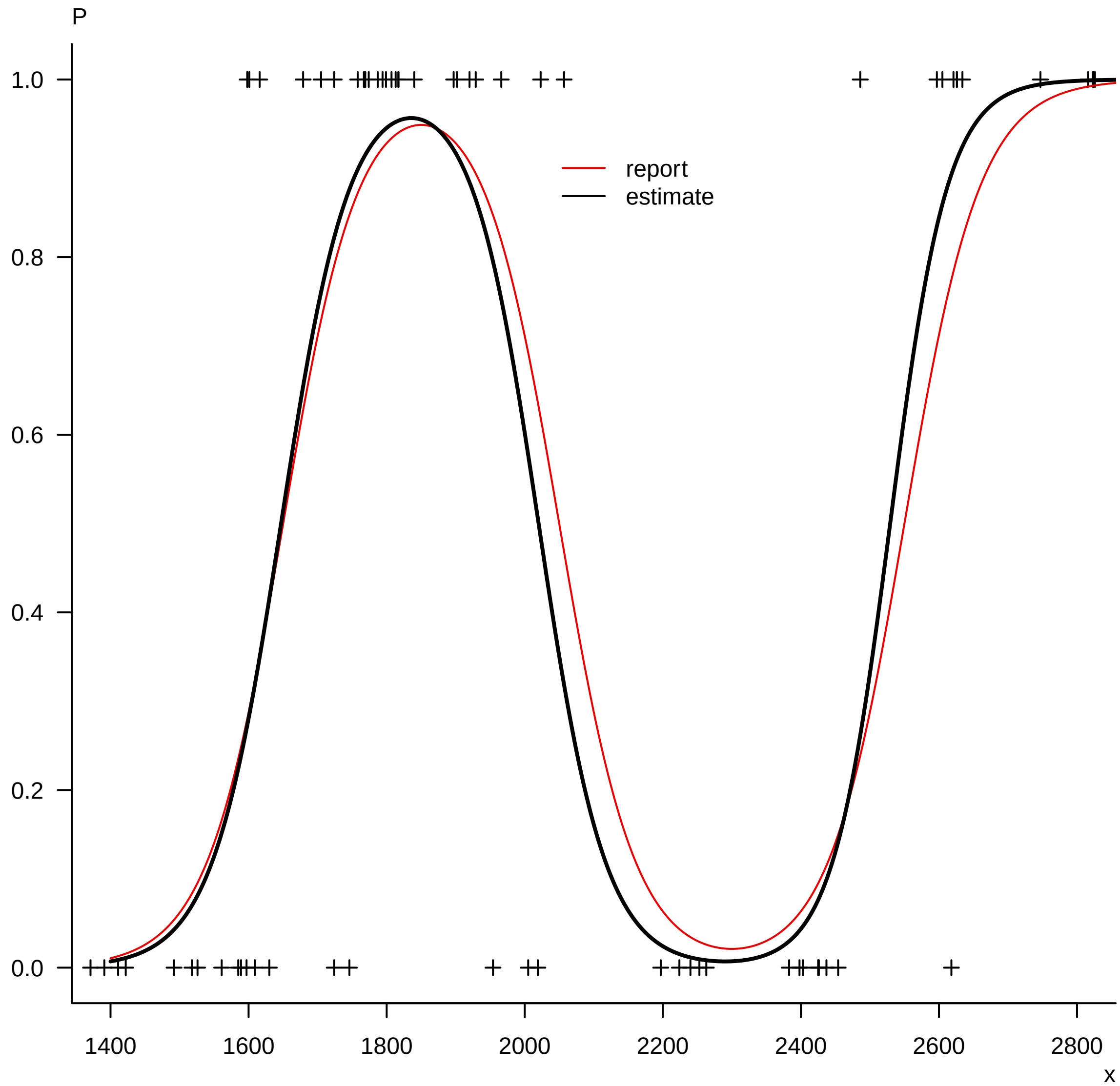

Fig. 1 Chang-Bodt model

4. Parametric QR Models

GLM with the Bernoulli response provides a way to characterize a possibly nonincreasing function as $P(x) = G\left(f(x)\right)$, where G is a monotone function such as the standard logistic, normal, or Cauchy CDF. Finite-dimensional parametric models are obtained by using some basis $(f_1, f_2, \ldots)$, and then $f(x) = X^t\beta$ for fixed k where $\beta = (\beta_1, \ldots, \beta_k)$ and $X = \left(f_1(x), \ldots, f_k(x)\right)$. So, we get

$$P(x) = G(X^t\beta)\,. \tag{10}$$

This accounts for the (monotonic) basic linear or location-scale model

$$P(x) = G(b_0 + b_1 x) = G\left(\frac{x-m}{s}\right) \tag{11}$$

and polynomials of arbitrary degree

$$P(x) = G\left(\sum_{i=0}^{k} b_i x^i\right)\,. \tag{12}$$

The canonical polynomial basis is given by $f_i(x) = x^i$. In practice, we use an orthogonal polynomial basis, with degree$(f_i) = i$ and $\int f_i f_j = 0$ if $i \neq j$. This eliminates numerical problems and provides the same solution as the canonical basis for each k. Other popular choices for basis sets include the natural spline and B-spline. See Fig. 2 for basis set examples with dimension $k = 5$ where various colors distinguish the basis elements. Figure 3 shows logistic response estimates for these 3 basis sets with dimension $d = 3, 4, 5, 6$.

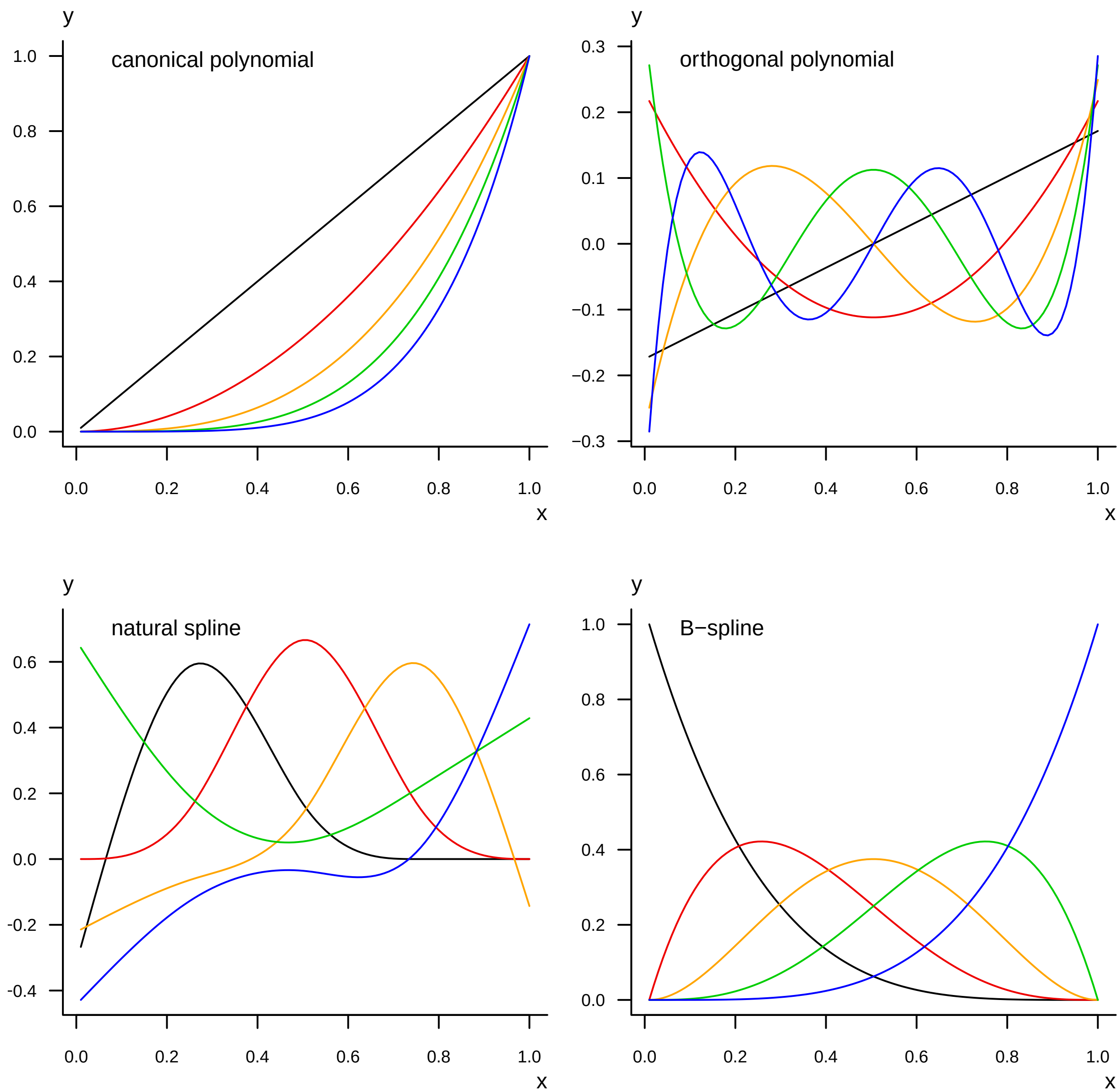

Fig. 2 Regression basis sets

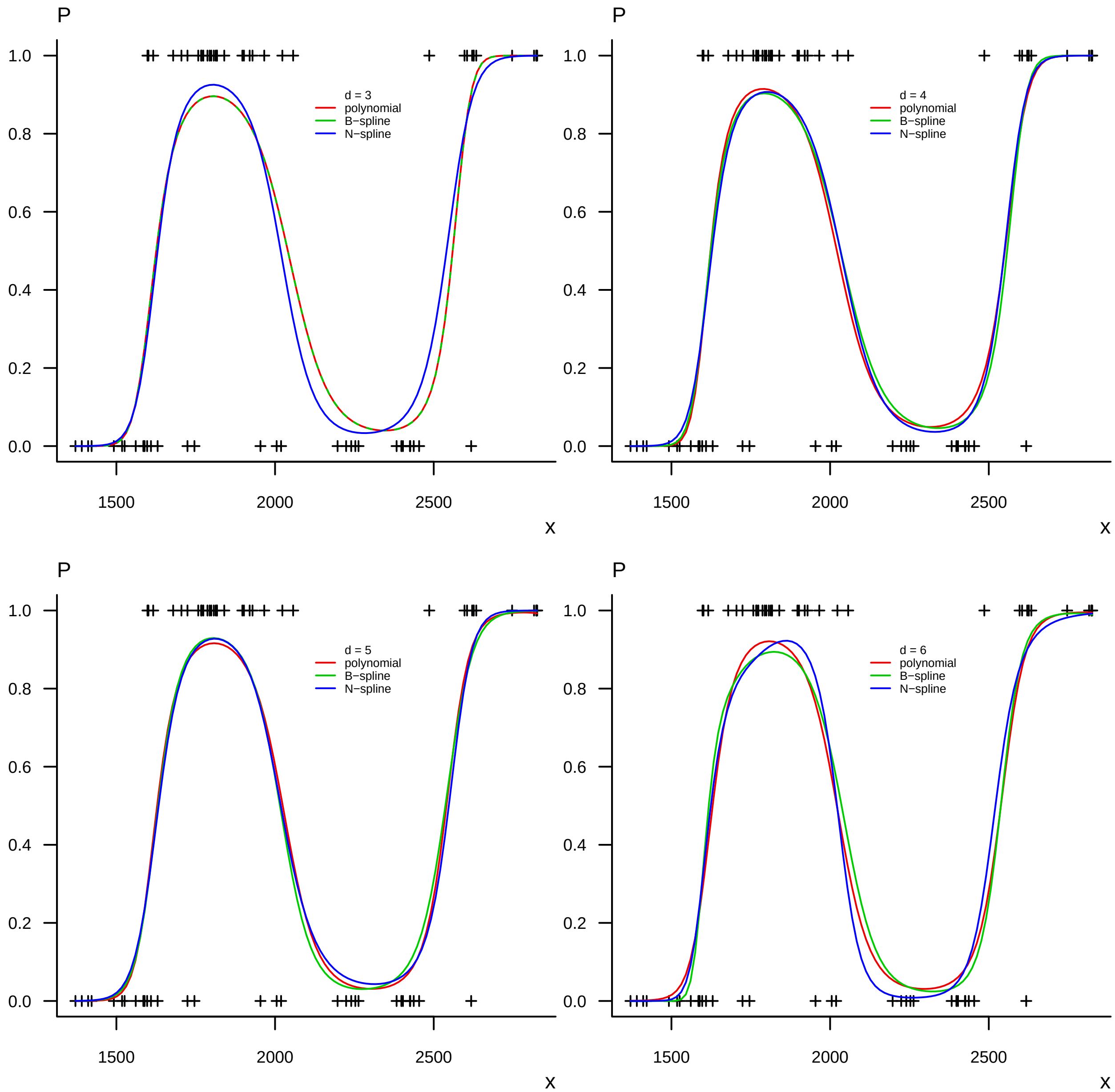

Fig. 3 Logistic regression

Approved for public release; distribution is unlimited.

5. Nonparametric QR Models

Nonparametric linear models are presented in Appendix A. These arise from a process that can be described as penalized maximum likelihood estimation, penalized least squares, or smoothing.

In Appendix B, the iteratively reweighted least squares (IRLS) GLM estimation procedure for QR models is adapted to use nonparametric penalized linear models. This gives rise to nonparametric penalized QR models.

5.1 The P-spline QR Model

Eilers and Marx[4] propose using an overfitted B-spline model and then penalizing to produce a smooth fit. This is called a P-spline model. In this application, we fit a penalized GLM with a degree 3 B-spline basis of size $p = 32$ penalized with the second derivative D. The smoothing operator is $S = e^{\lambda} D$.

Selection of the smoothing parameter is usually accomplished by optimizing some information or cross-validation quantity such as the Akaike information criterion (AIC), ordinary cross-validation (OVC), or generalized cross-validation (GCV) as described in Appendix A. There are other such quantities and other methods, and no single procedure is known to give the best solution. In fact, no single procedure even works for all data sets. So the choice of smoothing parameter selection procedure is itself somewhat subjective. To compensate for the perception that the procedures tend to oversmooth the response, an adjusted smoothing parameter is computed as the minimum of the 3 values obtained less their range. For this data, $\lambda_{aic} = -6.2$, $\lambda_{gcv} = -6.0$, $\lambda_{ocv} = -5.7$, and $\lambda_{adj} = -6.7$.

See Fig. 4 for fits with various values of the smoothing parameter. Optimal solutions for the 3 smoothing selection methods along with the adjusted solution are shown in Fig. 5. The adjusted solution and confidence intervals are shown in Fig. 6.

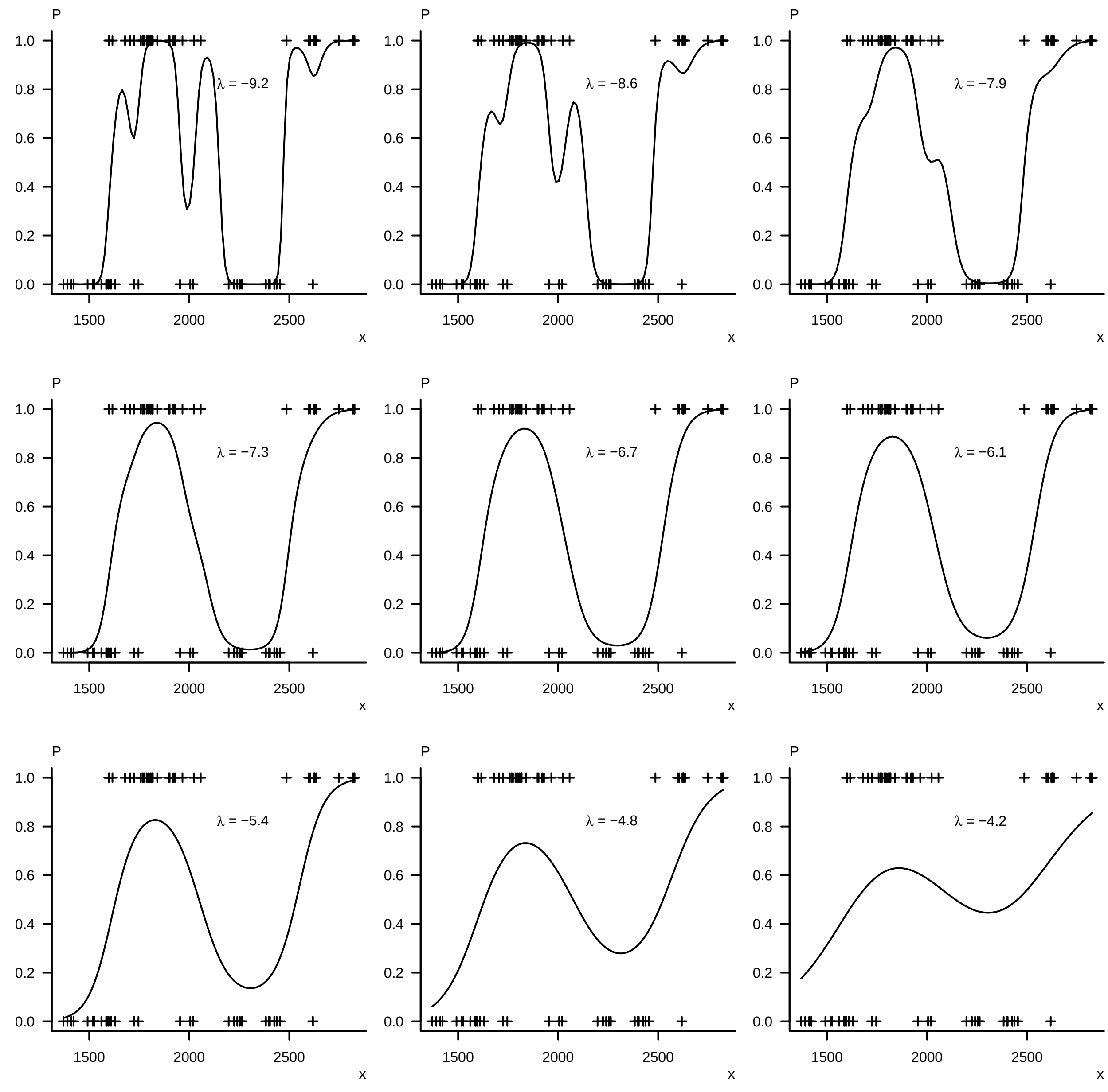

Fig. 4 P-spline optimization

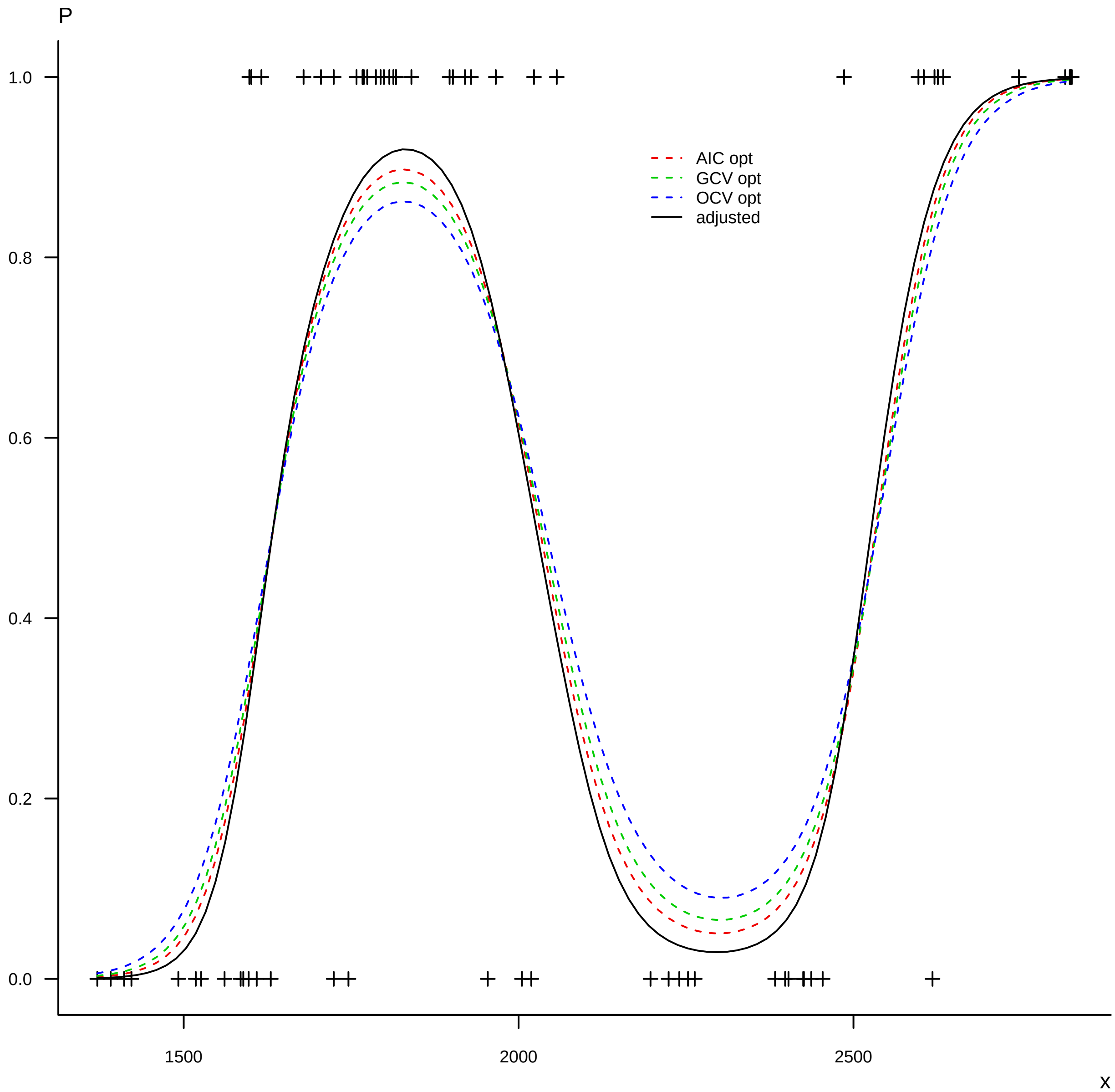

Fig. 5 Optimal and adjusted P-spline solutions

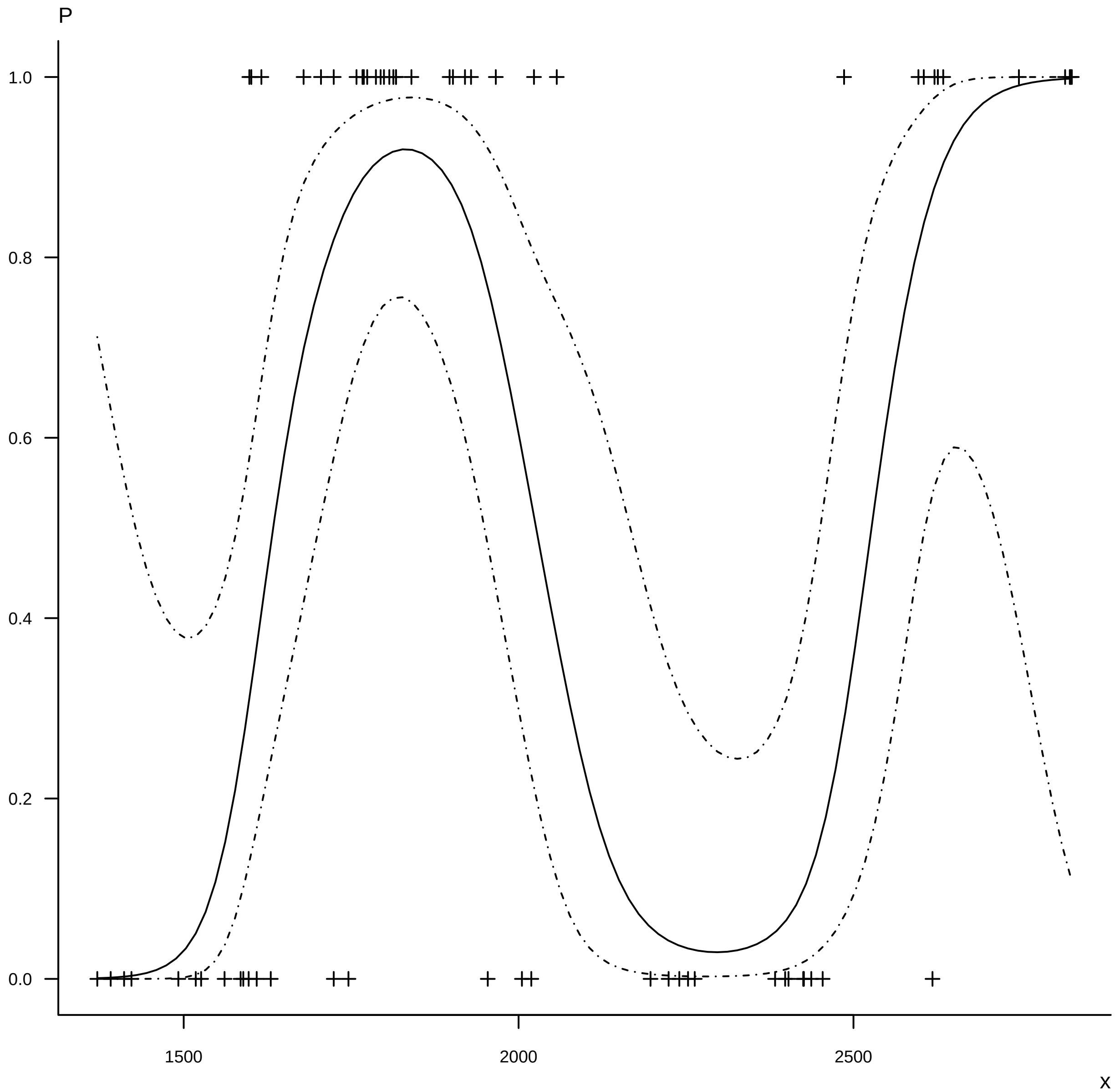

Fig. 6 Adjusted P-spline solution with 90% confidence intervals

5.2 The S-spline QR Model

The smoothing spline, or S-spline, QR model has the form $P(x) = G(f(x))$, where f is an S-spline linear model with knots at the stimulus data points, and G is an arbitrary link function. Wahba[5] is the standard reference for smoothing splines.

Smoothing splines are obtained through optimization in certain function spaces, and the GLM implementation described in Section B.3 accomplishes this using the standard IRLS GLM algorithm. In this application, we fit a cubic smoothing spline GLM by penalizing with the second derivative D. The smoothing operator is $S = e^{\lambda} D$.

We allow for multiple observations at a single stimulus by averaging the response and multiplying the weight by the observation multiplicity at that level. However, the sum of squared errors (SSE) and Λ are computed from the original data, so they are comparable with the other models.

Smoothing parameter selection methodology is the same as for the P-spline, Section 5.1. Typically, different λ values are obtained.

See Fig. 7 for fits with various values of the smoothing parameter. Optimal solutions for all 3 smoothing selection methods and the adjusted solution are shown in Fig. 8. The adjusted solution and confidence intervals are shown in Fig. 9.

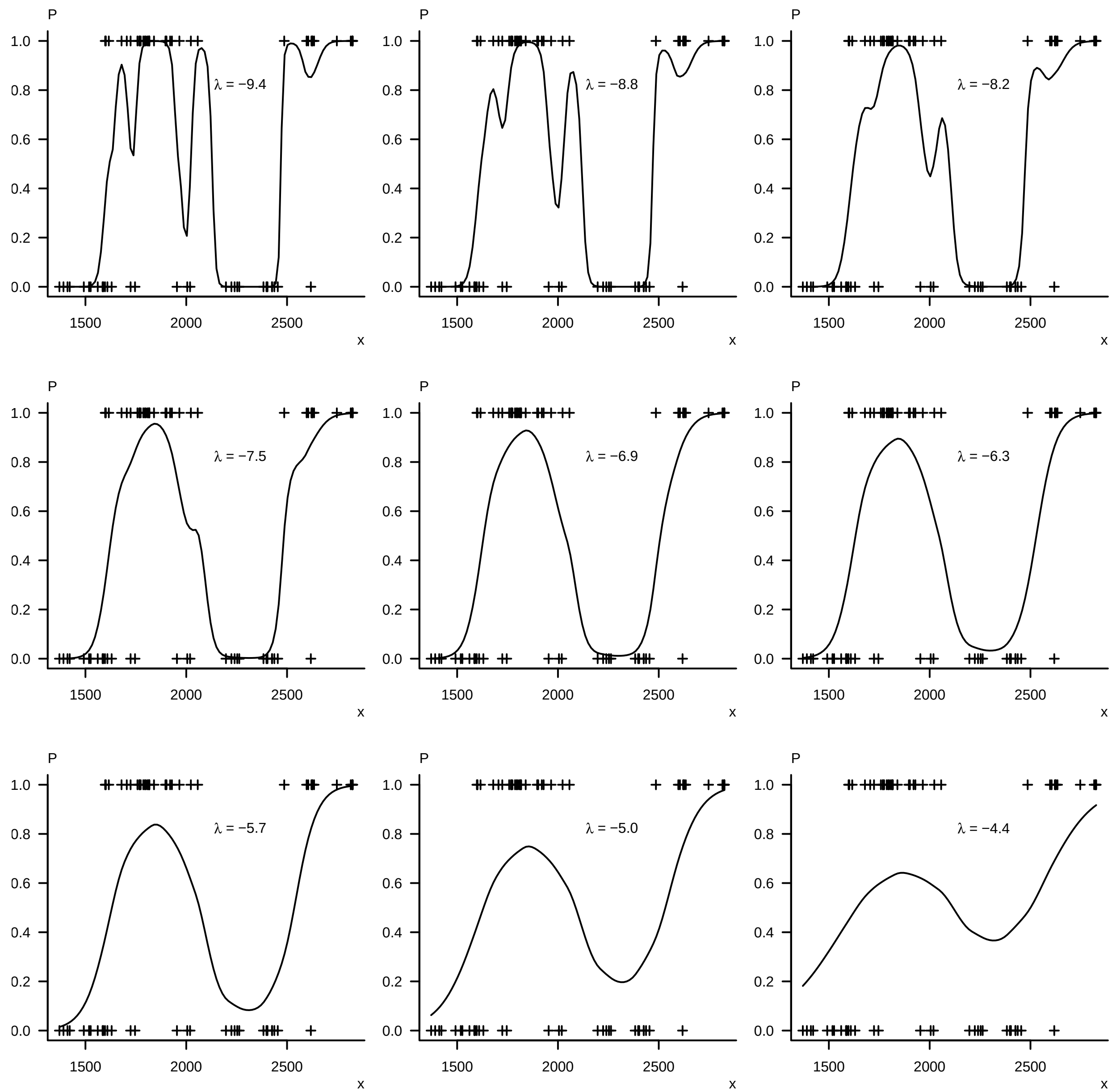

Fig. 7 S-spline optimization

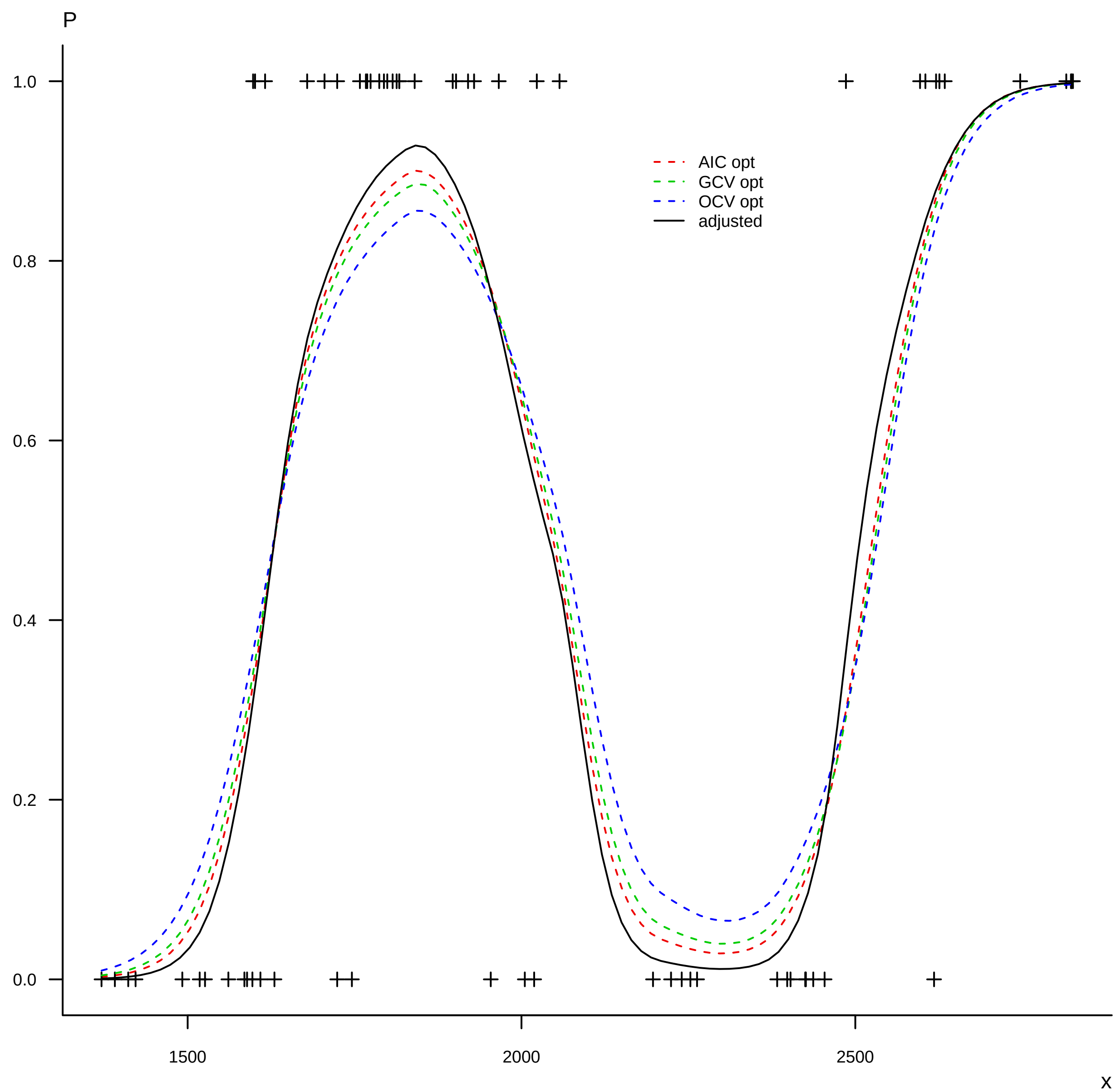

Fig. 8 Optimal S-spline solutions

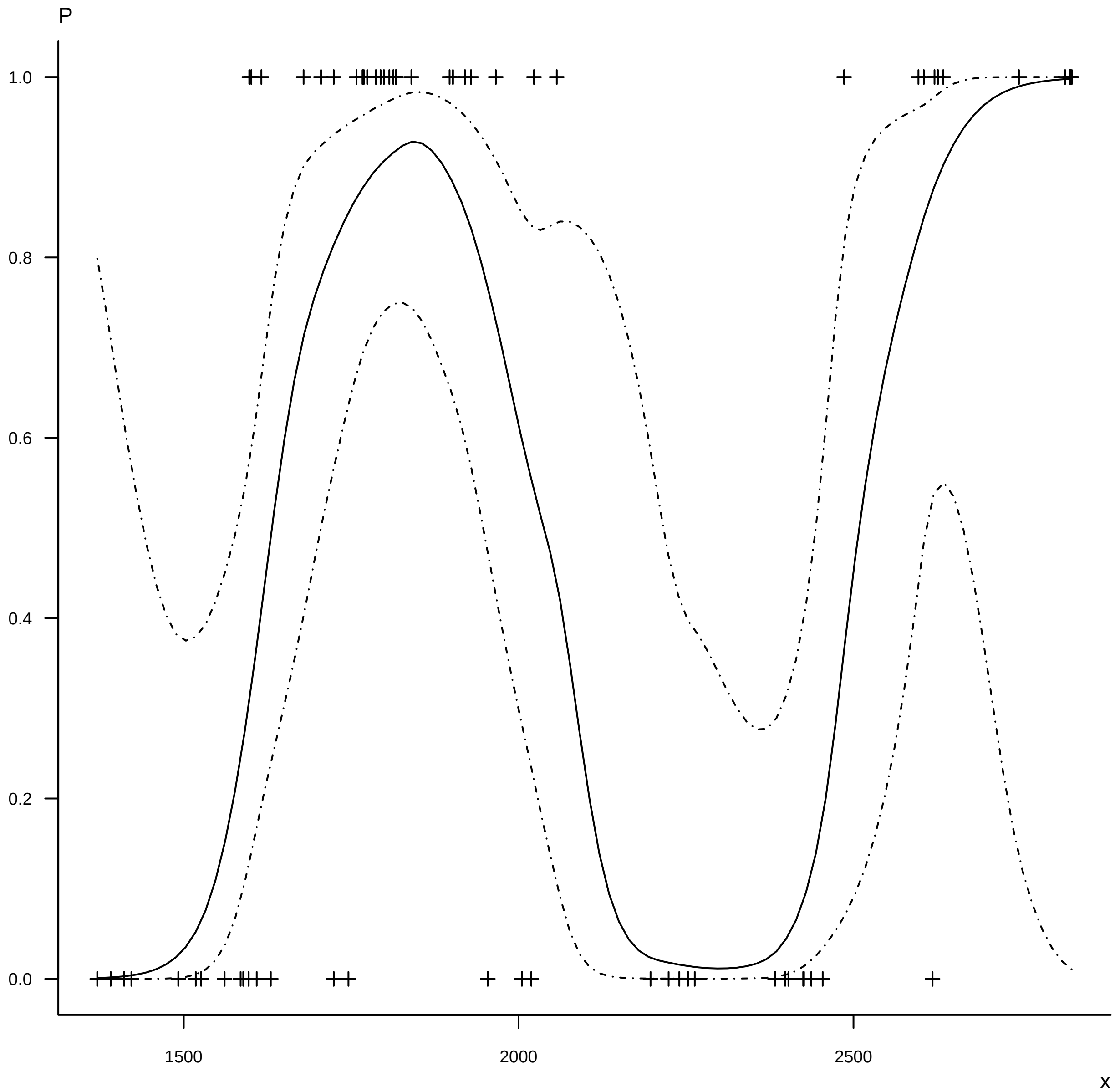

Fig. 9 Adjusted S-spline solution with 90% confidence intervals

6. Model Comparison

Table 2 displays fit statistics and Fig. 10 shows the superimposed solutions for selected methods: degree 6 parametric models, reported and estimated Chang-Bodt (CB), and adjusted P-spline and S-spline. The adjusted smoothing spline solution is the best fit in terms of squared error and maximum likelihood measures.

Table 2 Goodness of fit comparison

Model	SSE	Λ
Polynomial (6)	7.237	22.49
B-spline (6)	7.212	22.34
Natural spline (6)	7.201	22.07
CB reported	7.310	22.58
CB estimated	7.186	21.91
P-spline adjusted	7.045	21.98
S-spline adjusted	6.847	21.37

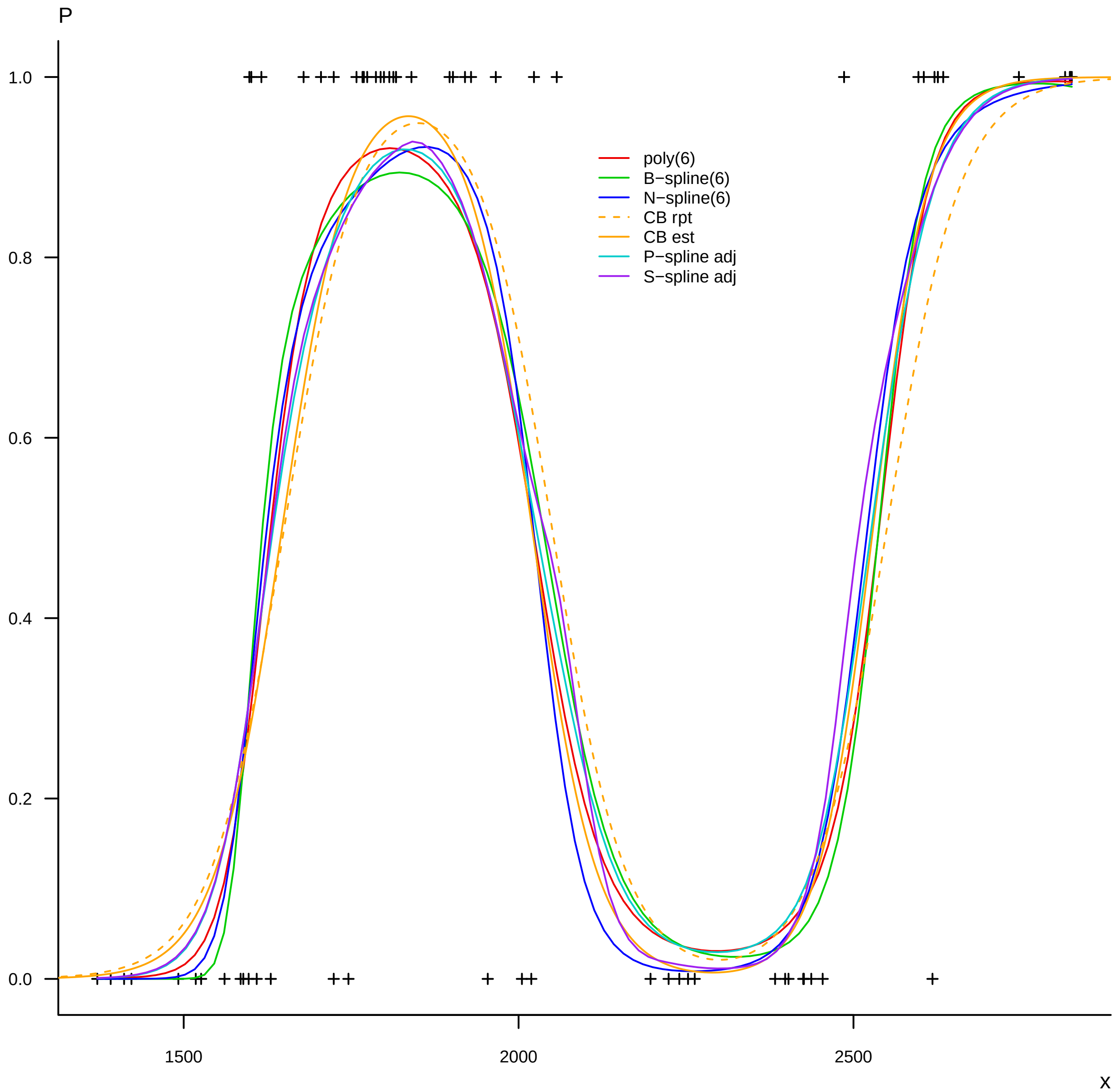

Fig. 10 All solutions

7. Conclusions and Recommendations

The simple location-scale quantal response model may not adequately capture the behavior of observed phenomena.

Higher-order polynomial and finite-dimensional spline basis models allow for more complicated responses as the polynomial degree or spline basis dimension increases.

Penalized B-spline (P-spline) and smoothing spline (S-spline) models offer the most flexibility as these are nonparametric (not constrained to any particular functional form). These should be useful in identifying nonstandard behavior via statistical goodness-of-fit tests.

8. References

1. Collins J. Quantal response: practical sensitivity testing. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2012 Jun. Report No.: ARL-TR-6022. 1

2. Collins J. Quantal response: estimation and inference. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2014 Sep. Report No.: ARL-TR-7088. 1

3. Chang AL, Bodt BA. JTCG/AS interlaboratory ballistic test program—final report. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 1997 Dec. Report No.: ARL-TR-1577. 2

4. Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. Statistical Science. 1996;11(2):89–121. 8

5. Wahba G. Spline models for observational data. Philadelphia (PA): Society for Industrial and Applied Mathematics; 1990. 12

INTENTIONALLY LEFT BLANK.

Appendix A. The Linear Model

A.1 The Basic Linear Model

The usual linear model is

$$Y = X\beta + \varepsilon\,, \tag{A-1}$$

where Y is an $n\times 1$ response, X is an $n\times p$ independent matrix, β is a $p\times 1$ parameter, and the $n\times 1$ error $\varepsilon \sim N(0,\sigma^2 I_n)$ is normally distributed.

For any positive-definite symmetric matrix W, the corresponding weighted inner product and norm are, respectively, $\langle x,y\rangle_W = x^t W y$ and $\|x\|_W = \langle x,x\rangle_W^{1/2}$. So, in general, the density of a normal vector with mean M and variance-covariance matrix V can be written as

$$f(x) = (2\pi)^{-n/2}|V|^{-1/2}\exp\left[-\tfrac{1}{2}\|x-M\|_{V^{-1}}^2\right]. \tag{A-2}$$

For the model of Eq. A-1, we have $\mathrm{E}\,Y = M = X\beta$ and $\mathrm{Var}\,Y = V = \sigma^2 I_n$. Each column of X is a linear predictor, and the model is

$$y_i = \sum_{j=1}^{p} X_{ij}\beta_j\,. \tag{A-3}$$

We can work with a p-parameter model for a single predictor v by choosing a set of fixed basis functions $\{f_1,\ldots,f_p\}$ and setting $X_{ij} = f_j(v_i)$.

$$y_i = \sum_{j=1}^{p} \beta_j f_j(v_i)\,. \tag{A-4}$$

For example, the choice of $f_j(v) = v^{j-1}$ gives the polynomial model

$$y = \beta_0 + \beta_1 v + \beta_2 v^2 + \cdots \beta_{p-1}v^{p-1}\,. \tag{A-5}$$

See Eq. A-2. Solution by least squares is equivalent to maximum likelihood estimation for normal error, and the criterion is to choose u that minimizes $Q = \varepsilon^t\varepsilon = \|\varepsilon\|^2 = \|Y - X\beta\|^2$ since $\varepsilon = Y - X\beta$. This is

$$Q = \|Y\|^2 - 2\beta^t X^t Y + \|X\beta\|^2\,, \tag{A-6}$$

and by setting the derivative to 0,

$$\frac{dQ}{d\beta} = -2X^tY + 2X^tX\beta = 0\,, \tag{A-7}$$

we obtain the normal equations

$$X^tX\beta = X^tY \tag{A-8}$$

with solution

$$\hat{\beta} = (X^tX)^{-1}X^tY \tag{A-9}$$

and response estimate

$$\hat{Y} = HY\,, \tag{A-10}$$

where the so-called hat matrix is

$$H = X(X^tX)^{-1}X^t\,. \tag{A-11}$$

Note that $\mathrm{E}\,\hat{\beta} = \beta$ and $\mathrm{Var}\,\hat{\beta} = \sigma^2(X^tX)^{-1}$.

A.2 The Weighted Model

When the error is $N(0,\Sigma)$, the correct inner product is weighted by the symmetric $W = \Sigma^{-1}$, and so $Q = \varepsilon^tW\varepsilon = \|\varepsilon\|_W^2 = \|Y - X\beta\|_W^2$. This is

$$Q = \|Y\|_W^2 - 2\beta^tX^tWY + \|X\beta\|_W^2\,. \tag{A-12}$$

Then

$$\frac{dQ}{d\beta} = -2X^tWY + 2X^tWX\beta = 0\,, \tag{A-13}$$

the normal equations are

$$X^tWX\beta = X^tWY\,, \tag{A-14}$$

the solution is

$$\beta = (X^tWX)^{-1}X^tWY\,, \tag{A-15}$$

and the response estimate is $\hat{Y} = HY$, where

$$H = X(X^tWX)^{-1}X^tW\,. \tag{A-16}$$

Note that $\mathrm{E}\,\hat{\beta} = \beta$ and $\mathrm{Var}\,\hat{\beta} = (X^tWX)^{-1}$.

For these models, the likelihood function is

$$L(\beta) = (2\pi)^{-n/2}|\Sigma|^{-1/2}\exp\Big[-\tfrac{1}{2}\|Y - X\beta\|^2_{\Sigma^{-1}}\Big]\,, \tag{A-17}$$

its log derivative is

$$\frac{d}{d\beta}\log L(\beta) = X'\Sigma^{-1}(Y - X\beta)\,, \tag{A-18}$$

and the information matrix is

$$M_\beta = X'\Sigma^{-1}X\,. \tag{A-19}$$

A.3 Smoothing

If the system is ill-conditioned, we can maximize a penalized likelihood function

$$L_S(\beta) = L(\beta)\cdot\exp[-\tfrac{1}{2}\|S\beta\|^2]\,. \tag{A-20}$$

Equivalently, penalization can be applied to the least-squares formulation by minimizing $Q = \|\varepsilon\|^2_W + \|S\beta\|^2$ for some smoothing operator S. This is

$$Q = \|Y\|^2_W - 2\beta^t X^t W Y + \|X\beta\|^2_W + \|\beta\|^2_{S^t S}\,. \tag{A-21}$$

Then

$$\frac{dQ}{d\beta} = -2X^t W Y + 2X^t W X\beta + 2S^t S\beta = 0\,, \tag{A-22}$$

and the normal equations are

$$(X^t W X + S^t S)\beta = X^t W Y \tag{A-23}$$

with solution

$$\beta = (X^t W X + S^t S)^{-1} X^t W Y \tag{A-24}$$

and response estimate $\hat{Y} = HY$ where

$$H = X(X^t W X + S^t S)^{-1} X^t W\,. \tag{A-25}$$

This is equivalent to the model $Y^* = X^*\beta + \varepsilon^*$ with weight W^* where $Y^* = \left[\begin{smallmatrix} Y \\ 0 \end{smallmatrix}\right]$,

$X^* = \left[\begin{smallmatrix} X \\ S \end{smallmatrix}\right]$, and $W^* = \left[\begin{smallmatrix} W & 0 \\ 0 & I \end{smallmatrix}\right]$. Then the normal equations are

$$\begin{aligned}
\begin{bmatrix} X^t & S^t \end{bmatrix} \begin{bmatrix} W & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} X \\ S \end{bmatrix} \beta &= \begin{bmatrix} X^t & S^t \end{bmatrix} \begin{bmatrix} W & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} Y \\ 0 \end{bmatrix} \\
\begin{bmatrix} X^t & S^t \end{bmatrix} \begin{bmatrix} WX \\ S \end{bmatrix} \beta &= \begin{bmatrix} X^t & S^t \end{bmatrix} \begin{bmatrix} WY \\ 0 \end{bmatrix} \\
(X^t W X + S^t S)\beta &= X^t W Y
\end{aligned} \tag{A-26}$$

as required. This so-called augmented representation is useful for computation.

Modern software for linear least-squares estimation operates on Eqs. A-8, A-14, and A-26 through the response vector Y, the weight vector W, the smoothing vector S, and the design matrix X. The normal equations are solved efficiently without inverting the design matrix, and we get parameter estimates and diagnostics such as the parameter variance and hat matrix diagonal.

A.4 Nonparametric P-spline

Smoothing or penalization can be applied to an overfitted parametric model as given by Eq. A-4. Suppose f has the particular form

$$f(x) = \sum_{j=1}^{p} \beta_j f_j(x_i) \tag{A-27}$$

for a B-spline basis $(f_1, \ldots, f_p)$ and $\mathcal{D}$ is a linear differential operator. Computational details for the B-spline are in Appendix C, and for the differential operator see Appendix D. The solution is the minimizer $(\beta_1, \ldots, \beta_p)$ of

$$\sum_{i=1}^{n} \Big(y_i - \sum_{j=1}^{p} \beta_j f_j(x_i) \Big)^2 + \lambda^2 \int_a^b \Big(\mathcal{D}\Big(\sum_{j=1}^{p} \beta_j f_j \Big)(u) \Big)^2 \, du \tag{A-28}$$

and is again given exactly by Eq. A-21. This works because of the ordered partial partition property of B-spline bases: smoothing the coefficient vector in fact smooths the solution. Eilers and Marx[1] call this a P-spline (penalized B-spline) model. Figure A-1 depicts the effect of smoothing parameter variation for the P-spline model.

[1]Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. Statistical Science. 1996;11(2):89–121.

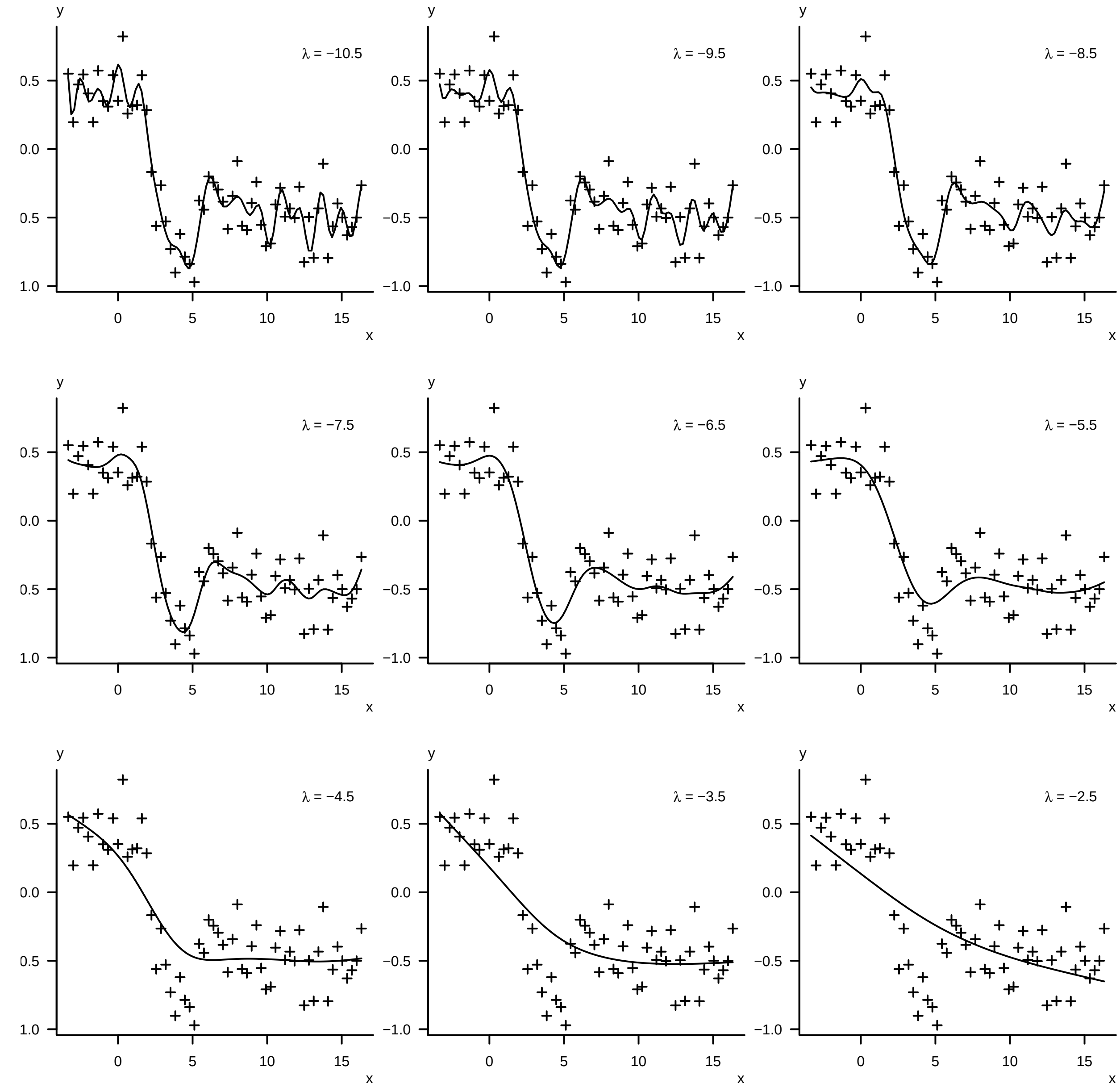

Fig. A-1 P-spline optimization, linear model

A.5 Nonparametric S-spline

The smoothing spline is another manifestation of penalized maximum likelihood or least squares estimation. A smoothing spline, or S-spline, f is the minimizer of

$$\sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda^2 \int_a^b ((\mathcal{D} f)(u))^2 \, du \,, \qquad \text{(A-29)}$$

where f is in some suitable function space, and $\mathcal{D}$ is a linear differential operator. The solution is a spline with knots at the data points. If $\mathcal{D}$ has order p, the solution is piecewise polynomial of degree $2p - 1$ with continuous derivatives of orders $0, 1, 2, \ldots, 2p - 2$. See Wahba[2].

The discrete representation of this problem is Eq. A-21, with $X = I$ and S taken to be a scalar multiple of a differential operator D, so $S = \lambda D$ and $S^t S = \lambda^2 D^t D$. The representation D of $\mathcal{D}$ is given in Appendix D. The discrete formulation provides the exact solution of Eq. A-29.

When $\mathcal{D}$ is the second derivative, the solution is a cubic spline. The solution is given at the data points, and spline interpolation can be used to evaluate the response at other values.

Figure A-2 depicts the effect of smoothing parameter variation for the S-spline model.

[2]Wahba G. Spline models for observational data. Philadelphia (PA): Society for Industrial and Applied Mathematics; 1990.

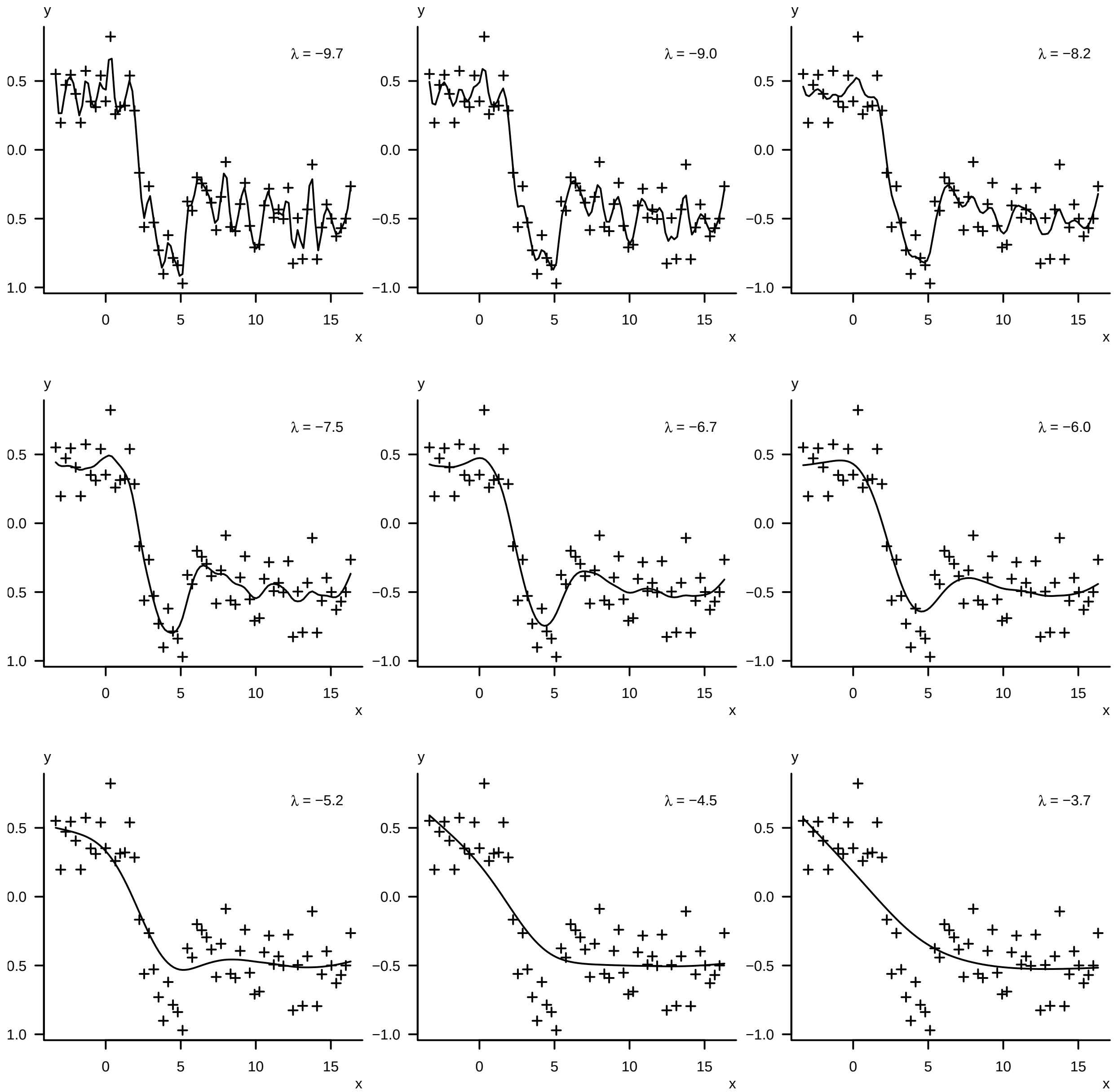

Fig. A-2 S-spline optimization, linear model

A.6 Selecting the Smoothing Parameter

Choice of smoothing parameter λ is usually based on minimization of one of the following quantities. The ordinary cross-validation (OCV) is

$$\mathrm{OCV} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{1 - H_{ii}}\right)^2 . \tag{A-30}$$

The generalized cross-validation (GCV) is

$$\mathrm{GCV} = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\left[\frac{1}{n}\operatorname{Tr}(I - H)\right]^2} . \tag{A-31}$$

The Akaike information criterion (AIC) is

$$\mathrm{AIC} = \Delta + 2 \cdot \operatorname{Tr} H , \tag{A-32}$$

where the deviance $\Delta = -2\log(L_{\text{reduced}}/L_{\text{full}})$ is defined in terms of the likelihood function L.

For normal error, $L_{\text{full}} = 1$, and based on Eq. A-2, we have

$$\log L_{\text{reduced}} = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}\|\varepsilon\|^2_{\Sigma^{-1}} , \tag{A-33}$$

where the residuals are $\varepsilon = Y - X\beta$ and the error variance is Σ.

For independent and identically distributed (IID) errors, $\Sigma = \sigma^2 I_n$ and $|\Sigma| = (\sigma^2)^n$. And so

$$\log L_{\text{reduced}} = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\varepsilon^t\varepsilon . \tag{A-34}$$

We replace σ^2 by the estimator $\varepsilon^t\varepsilon/n$ and get

$$\log L_{\text{reduced}} = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\frac{\varepsilon^t\varepsilon}{n}\right) - \frac{n}{2} = -\frac{n}{2}\left[\log\left(2\pi\frac{\varepsilon^t\varepsilon}{n}\right) + 1\right] \tag{A-35}$$

so

$$\Delta = n\left[\log\left(2\pi\frac{\varepsilon^t\varepsilon}{n}\right) + 1\right] . \tag{A-36}$$

See Fig. A-3 for graphs of OCV, GCV, and AIC as functions of λ. Note that AIC optimization fails for the S-spline, as no minimum value is obtained. See Fig. A-4 for GVC-optimal P-spline and S-spline solutions. With a maximum y-difference of less than 0.01, the solutions are practically indistinguishable.

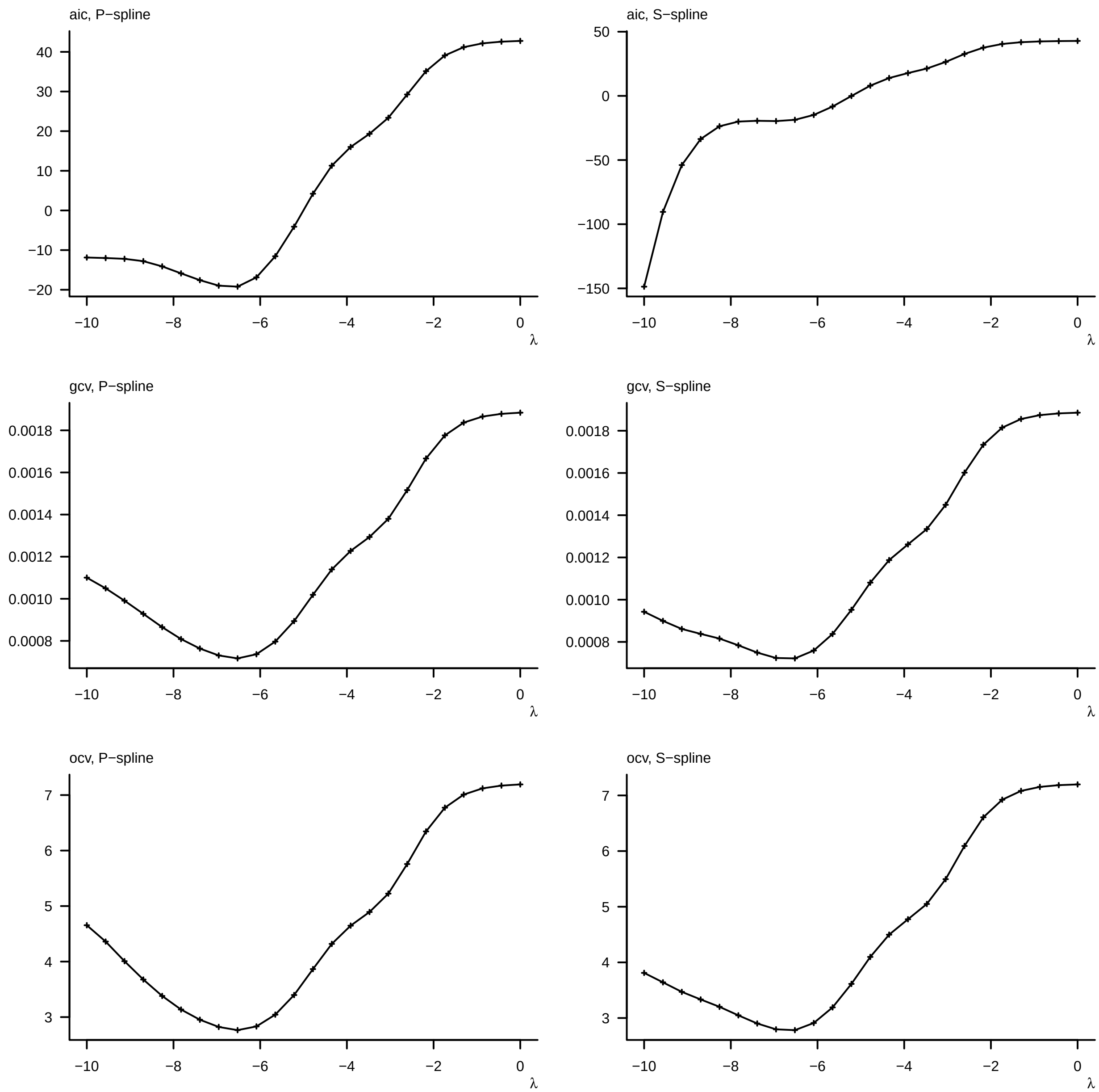

Fig. A-3 Smoothing parameter selection, linear model

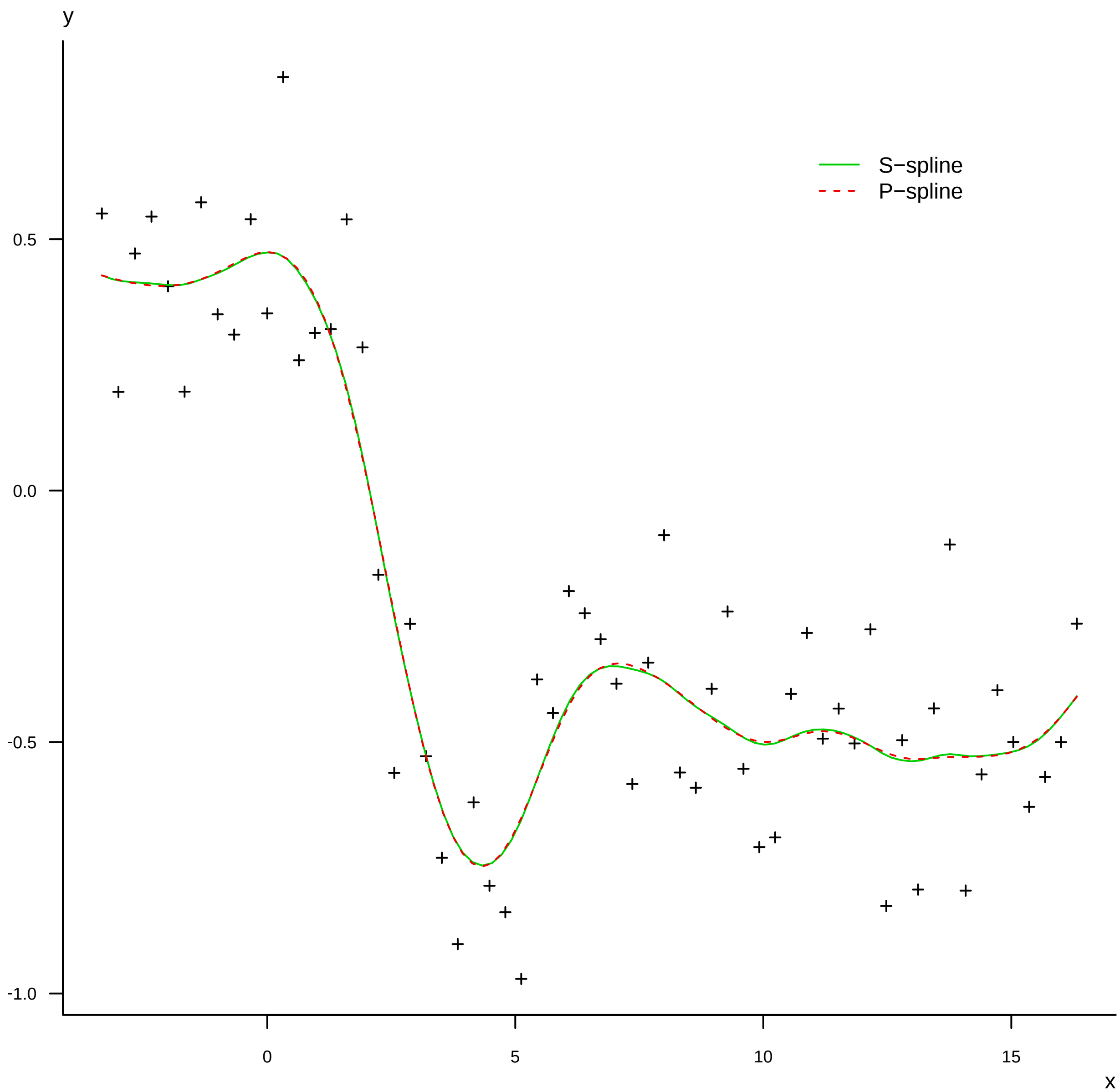

Fig. A-4 GCV-optimal fits, linear model

INTENTIONALLY LEFT BLANK.

Appendix B. The Generalized Linear Model

B.1 Generalized Linear Model Formulation

In the Generalized Linear Model (GLM), the response Y has an arbitrary distribution, $\eta = X\beta$ is a linear function of the parameter β, and the mean response is modeled as

$$\mu = \mathrm{E}[\, Y \mid X \,] = G(\eta) \tag{B-1}$$

for some monotone link function G with derivative $g = G'$. (Some authors call G^{-1} the link.) Response distributions are taken to be from a single-parameter exponential family, with the form

$$f(y,\theta,\psi) = \exp\left[\frac{y\theta - b(\theta)}{a(\psi)} + c(y,\psi)\right]. \tag{B-2}$$

The parameter θ is to be estimated, and ψ is a nuisance parameter.

With $\ell = \log f$, we calculate the moments of Y in terms of exponential family components.

$$\mathrm{E}\left[\frac{d}{d\theta}\ell(y,\theta,\psi)\right] = 0 \tag{B-3}$$

because $\int f(y,\theta,\psi)\,dy = 1$, and under suitable regularity conditions $\frac{d}{d\theta}\int f(y,\theta,\psi)\,dy$ $= \int \frac{d}{d\theta} f(y,\theta,\psi)\,dy = \int \frac{d}{d\theta}\ell(y,\theta,\psi)\cdot f(y,\theta,\psi)\,dy = 0$. Therefore, $E[(y - b'(\theta))/a(\psi)]$ $= 0$ and

$$E[Y] = \mu = G(\eta) = b'(\theta)\,. \tag{B-4}$$

Also, as usual,

$$\mathrm{E}\left[\left(\frac{d}{d\theta}\ell(y,\theta,\psi)\right)^2\right] = -\,\mathrm{E}\left[\frac{d^2}{d\theta^2}\ell(y,\theta,\psi)\right] \tag{B-5}$$

because $\mathrm{Var}\left[\frac{d}{d\theta}\ell\right] = \mathrm{E}\left[(\frac{d}{d\theta}\ell)^2\right] = \int (\frac{d}{d\theta}\ell)^2 \cdot f\,dy = \int \frac{d}{d\theta}\ell \cdot \frac{d}{d\theta} f\,dy$

$= \frac{d}{d\theta}\ell \cdot f|_{-\infty}^{\infty} - \int \frac{d^2}{d\theta^2}\ell \cdot f\,dy = -\,\mathrm{E}\left[\frac{d^2}{d\theta^2}\ell\right]$.

Therefore, $\mathrm{E}[(y-\mu)^2/a(\psi)^2] = -\,\mathrm{E}[-b''(\theta)/a(\psi)]$ and

$$\mathrm{Var}[Y] = v(\mu)a(\psi) = b''(\theta)a(\psi) \tag{B-6}$$

where $v(\mu) = \mathrm{Var}[Y]/a(\psi) = b''(\theta) = g(\eta)\frac{d\eta}{d\theta}$.

In the case that $\eta = \theta$, and hence $\mu = G(\eta) = G(\theta)$, we say that G is the canonical link function. Then $\mu = b'(\theta) = G(\theta) = G(\eta)$ and $v(\mu) = b''(\theta) = g(\theta) = g(\eta)$.

B.2 Estimation

Let $[x_1, \ldots, x_n] = X^t$, so the (column) vector x_i is row i of X, $\eta_i = x_i^t \beta$, and $\theta_i = \theta(\eta_i)$. Maximum likelihood estimation for the GLM is accomplished by maximizing the log-likelihood function

$$\mathcal{L} = \sum_{i=1}^{n} \ell(y_i, \theta_i, \psi) = \sum_{i=1}^{n} \left[\frac{y_i \theta_i - b(\theta_i)}{a(\psi)} + c(y_i, \psi) \right] . \tag{B-7}$$

This is a weighted least squares problem where the design and weight depend on the unknown parameter, and it can be solved iteratively by the Newton-Raphson method.

B.2.1 The Newton-Raphson Method

In 1 dimension, we find a zero of F by linearizing and updating the current argument x_o to x by solving $F(x_o) + (x - x_o)F'(x_o) = 0$ to get $x = x_o - F(x_o)/F'(x_o)$. We optimize F by setting $F' = 0$, so the update is $x = x_o - F'(x_o)/F''(x_o)$.

The vector version is $x = x_o - [\frac{d^2}{dxx^t} F(x_o)]^{-1} \frac{d}{dx} F(x_o)$. If we take $x = x_o + \delta$, the increment $\delta = x - x_o$ satisfies $[\frac{d^2}{dxx^t} F(x_o)]\delta = -\frac{d}{dx} F(x_o)$. So we need some derivatives.

B.2.2 Gradient

Differentiating, we have the gradient (vector of first derivatives)

$$\mathcal{D}(\beta) = \frac{d\mathcal{L}}{d\beta} = a(\psi)^{-1} \cdot \sum_{i=1}^{n} [y_i - b'(\theta_i)] \cdot \frac{d\theta_i}{d\beta} . \tag{B-8}$$

Since $\frac{d}{d\beta} b'(\theta_i) = b''(\theta_i) \frac{d\theta_i}{d\beta} = v(\mu_i) \frac{d\theta_i}{d\beta}$ and $\frac{d}{d\beta} b'(\theta_i) = \frac{d\mu_i}{d\beta} = \frac{d}{d\beta} G(\eta_i) = \frac{d}{d\beta} G(x_i^t \beta) = g(\eta_i) x_i$, we get

$$\frac{d\theta_i}{d\beta} = \frac{g(\eta_i)}{v(\mu_i)} x_i . \tag{B-9}$$

So the gradient is

$$\mathcal{D}(\beta) = a(\psi)^{-1} \cdot \sum_{i=1}^{n} (y_i - \mu_i) \cdot \frac{g(\eta_i)}{v(\mu_i)} x_i = a(\psi)^{-1} \cdot X^t W_F U_F , \tag{B-10}$$

where, for $i = 1, \ldots, n$, the diagonal weight matrix W_F has elements $w_{Fi} = g(\eta_i)^2 / v(\mu_i)$

$$W_F = \operatorname{diag}\left[\frac{g(\eta_1)^2}{v(\mu_1)}, \cdots, \frac{g(\eta_n)^2}{v(\mu_n)}\right] \tag{B-11}$$

and the centered/scaled response vector U_F has elements $u_{Fi} = (y_i - \mu_i)/g(\eta_i)$

$$U_F = \left[\frac{y_1 - \mu_1}{g(\eta_1)}, \cdots, \frac{y_n - \mu_n}{g(\eta_n)}\right]. \tag{B-12}$$

B.2.3 Hessian

Using $\frac{d}{d\beta} v(\mu_i) = v'(\mu_i) \frac{d}{d\beta} G(x_i^t \beta) = v'(\mu_i) g(x_i^t \beta) x_i$,

$$\frac{d^2}{d\beta\beta^t}\theta_i = \frac{g'(\eta_i)v(\mu_i) - g(\eta_i)^2 v'(\mu_i)}{v(\mu_i)^2} x_i x_i^t, \tag{B-13}$$

and the Hessian (matrix of second derivatives) is

$$\begin{aligned}
\mathcal{H}(\beta) = \frac{d^2}{d\beta\beta^t}\mathcal{L} &= a(\psi)^{-1} \cdot \sum_{i=1}^n \left[-b''(\theta_i)\frac{d\theta_i}{d\beta}\frac{d\theta_i}{d\beta}^t + (y_i - \mu_i)\frac{d^2}{d\beta\beta^t}\theta_i\right] \\
&= a(\psi)^{-1} \cdot \sum_{i=1}^n \left[-\frac{g(\eta_i)^2}{v(\mu_i)} + (y_i - \mu_i)\frac{g'(\eta_i)v(\mu_i) - g(\eta_i)^2 v'(\mu_i)}{v(\mu_i)^2}\right] x_i x_i^t \\
&= -a(\psi)^{-1} \cdot X^t W_N X
\end{aligned} \tag{B-14}$$

where

$$W_N = W_F - W_D \tag{B-15}$$

and W_D is a diagonal matrix with diagonal elements

$$w_{Di} = (y_i - \mu_i)\frac{g'(\eta_i)v(\mu_i) - g(\eta_i)^2 v'(\mu_i)}{v(\mu_i)^2}. \tag{B-16}$$

Since $\mathrm{E}[W_D] = 0$, the expected value of the Hessian is

$$\mathrm{E}\,\mathcal{H}(\beta) = -a(\psi)^{-1} \cdot X^t W_F X. \tag{B-17}$$

The Fisher Information Matrix is

$$M_\beta = \mathrm{E}\left[\frac{d}{d\beta}\mathcal{L} \cdot \frac{d}{d\beta}\mathcal{L}^t\right] = -\mathrm{E}\left[\frac{d^2}{d\beta\beta^t}\mathcal{L}\right] = -\mathrm{E}\,\mathcal{H}(\beta) = a(\psi)^{-1} \cdot X^t W_F X, \tag{B-18}$$

and the asymptotic estimator distribution is $N\left(\beta, a(\psi) \cdot (X' W_F X)^{-1}\right)$.

B.2.4 Newton-Raphson

Now, apply the Newton-Raphson algorithm to iteratively solve the optimization.

For GLM, the Newton-Raphson update is $\beta = \beta_o + \delta$ where

$$\begin{aligned} \mathcal{H}(\beta_o)\delta &= -\mathcal{D}(\beta_o) \\ (X^t W_N X)\delta &= X^t W_F U_F \\ (X^t W_N X)\delta &= X^t W_N U_N \end{aligned} \tag{B-19}$$

with

$$U_N = W_N^{-1} W_F U_F \; . \tag{B-20}$$

These are the normal equations for minimization of $Q = \|U_N - X\delta\|^2_{W_N}$. Both W_N and U_N depend on β_o. The normal equations can be solved iteratively with an initial guess β_o by calculating $\eta = X\beta_o$, $\mu = G(\eta)$, g, g', v, v', W_F, U_F, W_D, W_N, and U_N. Then solve for δ. The updated solution is $\beta = \beta_o + \delta$. Now replace β_o with β, and repeat. This is iteratively reweighted least squares with the Newton-Raphson update.

B.2.5 Fisher Scoring

For the GLM, the Fisher scoring update uses $\mathrm{E}\mathcal{H}$ in place of $\mathcal{H}$ to get

$$\begin{aligned} \mathrm{E}\mathcal{H}(\beta_o)\delta &= -\mathcal{D}(\beta_o) \\ (X^t W_F X)\delta &= X^t W_F U_F \; , \end{aligned} \tag{B-21}$$

which are the normal equations for minimization of $Q = \|U_F - X\delta\|^2_{W_F}$. Both W_F and U_F depend on β_o. The normal equations can be solved iteratively with an initial guess β_o by calculating $\eta = X\beta_o$, $\mu = G(\eta)$, g, v, W_F, and U_F. Then solve for δ, update, and repeat as above. This is iteratively reweighted least squares with Fisher scoring.

B.3 Smoothing

Smoothing is accomplished by maximizing a penalized objective

$$\mathcal{L}_S = \mathcal{L} - \tfrac{1}{2}\|S\beta\|^2 \tag{B-22}$$

with gradient

$$\mathcal{D}_S(\beta) = \frac{d\mathcal{L}_S}{d\beta} = X^t W_F U_F - S^t S\beta\,, \tag{B-23}$$

Hessian matrix

$$\mathcal{H}_S(\beta) = \frac{d^2}{d\beta\beta^t}\mathcal{L}_S = -X^t W_N X - S^t S\,, \tag{B-24}$$

and expected Hessian

$$\mathrm{E}\,\mathcal{H}_S(\beta) = -X^t W_F X - S^t S\,. \tag{B-25}$$

The Newton-Raphson update equations are

$$\begin{aligned}
\mathcal{H}_S(\beta_o)\delta &= -\mathcal{D}_S(\beta_o)\\
(X^t W_N X + S^t S)(\beta - \beta_o) &= X^t W_F U_F - S^t S\beta_o\\
(X^t W_N X + S^t S)\beta &= (X^t W_N X + S^t S)\beta_o + X^t W_F U_F - S^t S\beta_o\\
\cdot \qquad &= X^t W_N X\beta_o + X^t W_F U_F\\
\cdot \qquad &= X^t W_N X\beta_o + X^t W_N W_N^{-1} W_F U_F\\
\cdot \qquad &= X^t W_N (X\beta_o + W_N^{-1} W_F U_F)\\
\cdot \qquad &= X^t W_N (X\beta_o + U_N)\\
(X^t W_N X + S^t S)\beta &= X^t W_N Z_N\,,
\end{aligned} \tag{B-26}$$

where

$$Z_N = X\beta_o + U_N\,. \tag{B-27}$$

These are the normal equations for minimization of $Q = \|Z_N - X\beta\|^2_{W_N} + \|S\beta\|^2$. The update increment δ is now implicit. Replace β_o with β and repeat.

The Fisher update equations are $\mathrm{E}\,\mathcal{H}_S(\beta_o)\beta = \mathrm{E}\,\mathcal{H}_S(\beta_o)\beta_o - \mathcal{D}_S(\beta_o)$, or

$$\begin{aligned}
\mathrm{E}\,\mathcal{H}_S(\beta_o)\delta &= -\mathcal{D}_S(\beta_o)\\
(X^t W_F X + S^t S)(\beta - \beta_o) &= X^t W_F U_F - S^t S\beta_o\\
(X^t W_F X + S^t S)\beta &= (X^t W_F X + S^t S)\beta_o + X^t W_F U_F - S^t S\beta_o\\
\cdot \qquad &= X^t W_F X\beta_o + X^t W_F U_F\\
\cdot \qquad &= X^t W_F (X\beta_o + U_F)\\
(X^t W_F X + S^t S)\beta &= X^t W_F Z_F\,,
\end{aligned} \tag{B-28}$$

where

$$Z_F = X\beta_o + U_F \, . \tag{B-29}$$

These are the normal equations for minimization of $Q = \|Z_F - X\beta\|^2_{W_F} + \|S\beta\|^2$.

At each step, the normal equations are solved as the augmented system of Section A.3. P-spline models are obtained with a B-spline basis as in Section A.4, and smoothing spline models are obtained with $X = I$ as in Section A.5.

Figure B-1 depicts the effect of smoothing parameter variation for the P-spline model, and Fig. B-2 depicts the effect of smoothing parameter variation for the S-spline model. See Fig. B-3 for graphs of OCV, GCV, and AIC as functions of λ, and see Fig. B-4 for GVC-optimal P-spline and S-spline solutions.

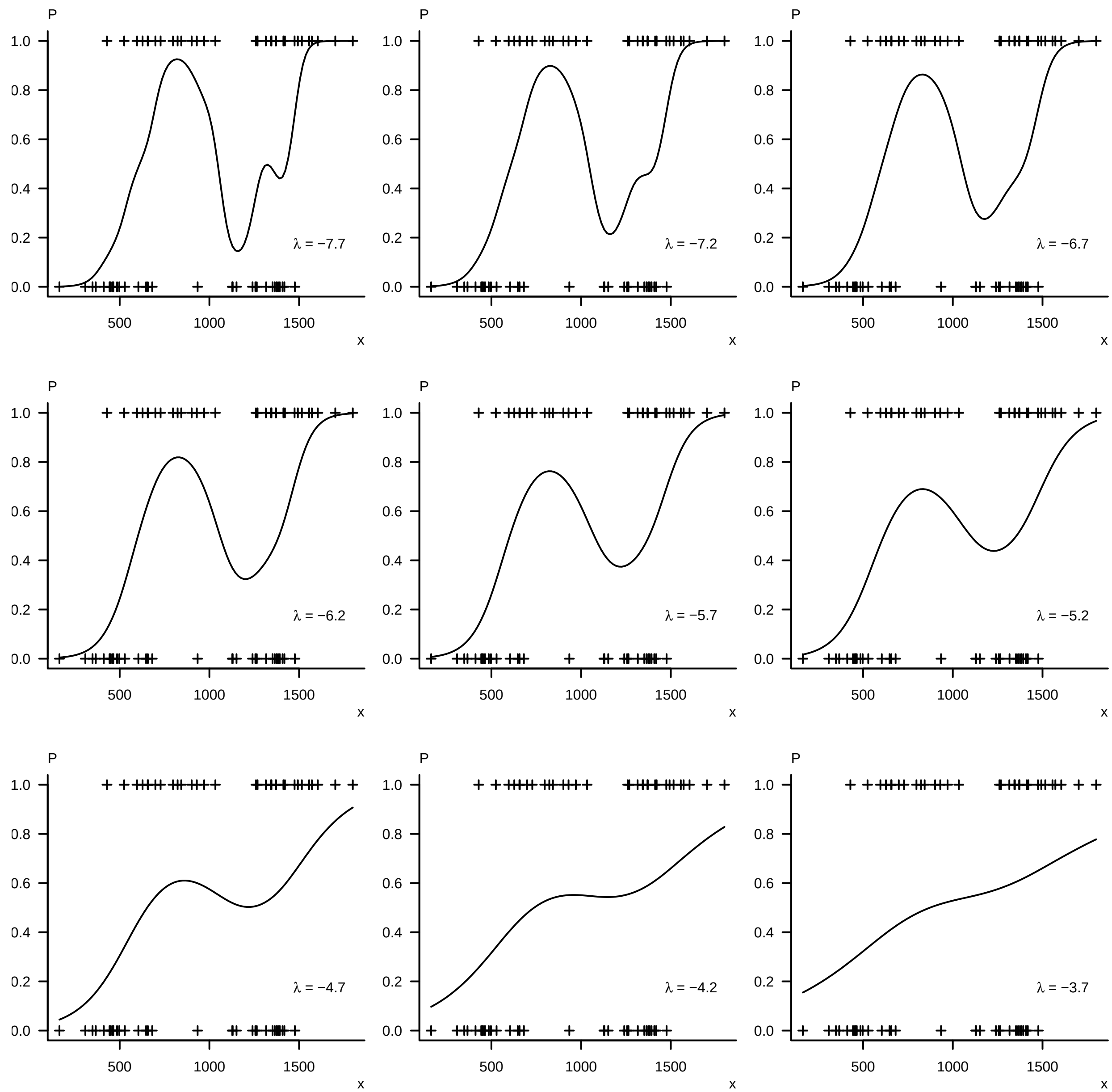

Fig. B-1 P-spline optimization, GLM

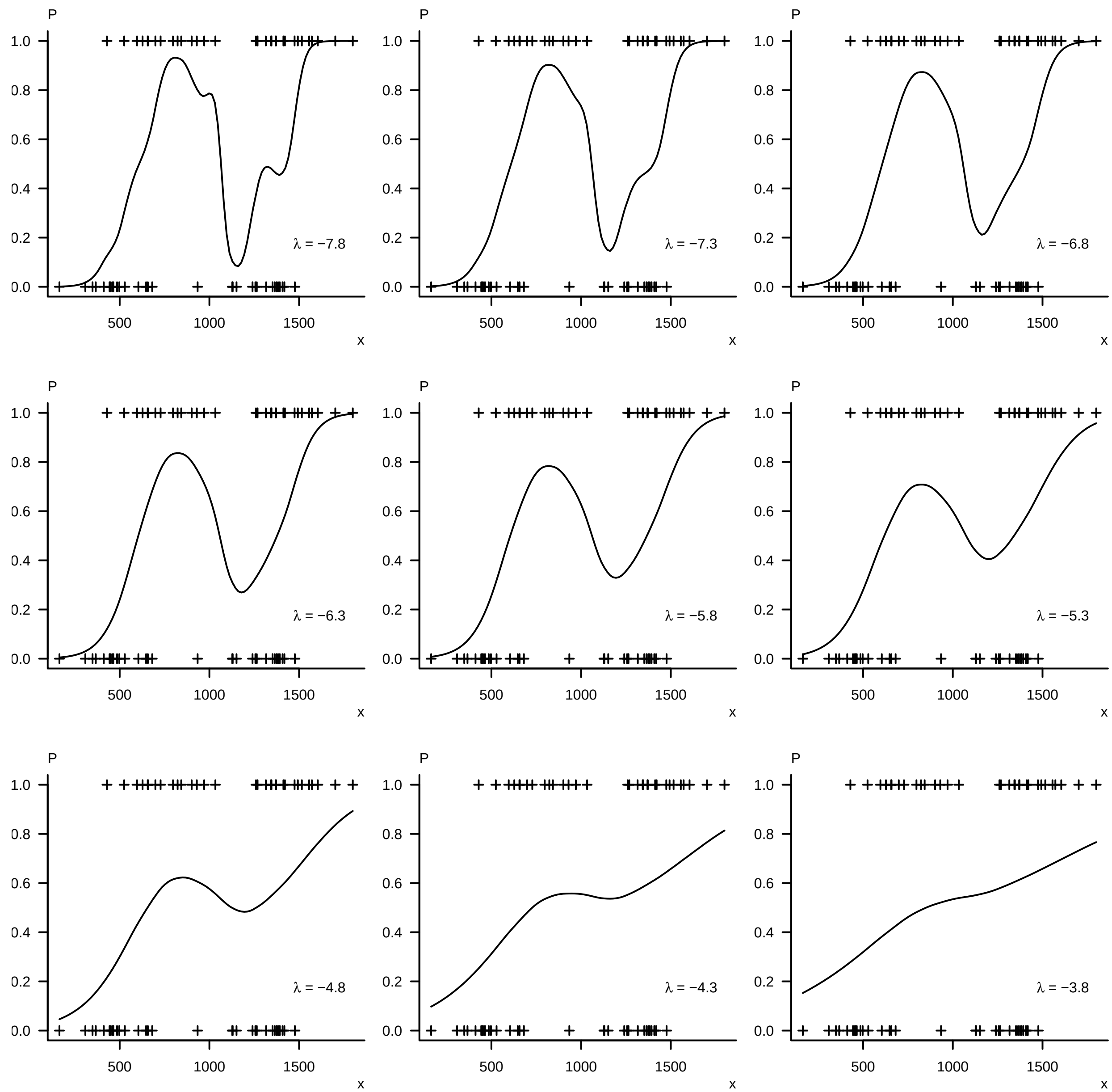

Fig. B-2 S-spline optimization, GLM

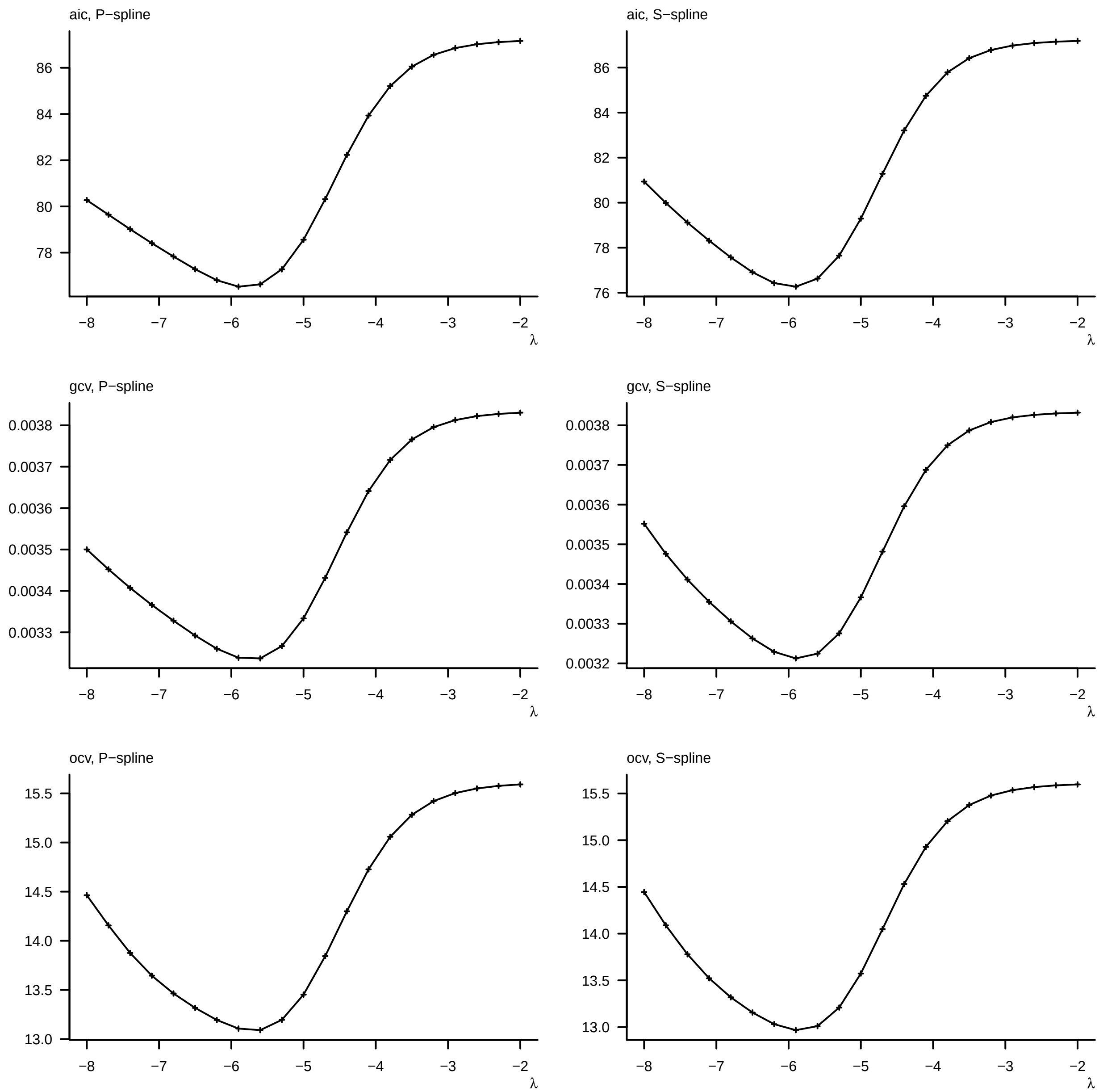

Fig. B-3 Smoothing parameter selection, GLM

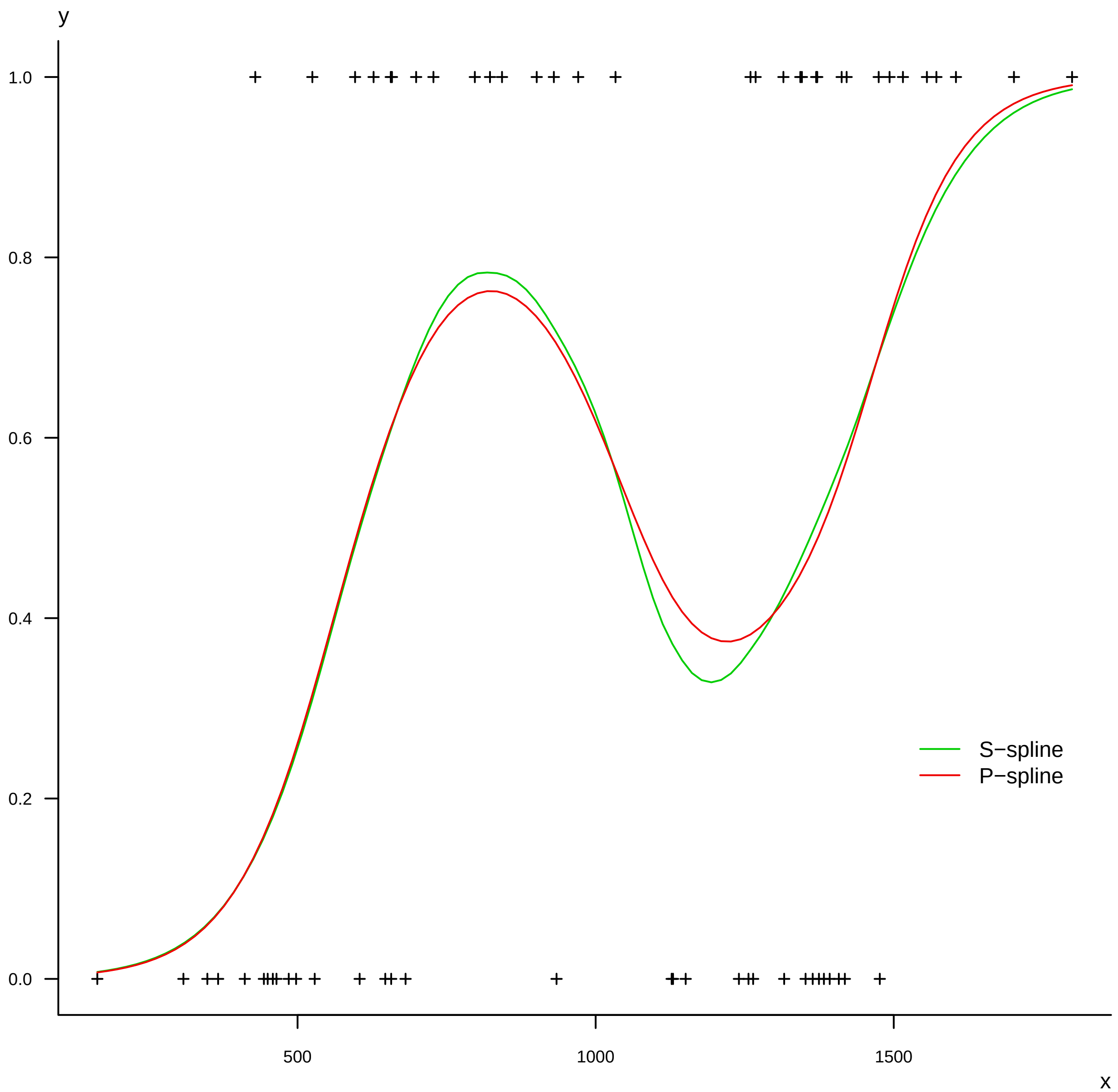

Fig. B-4 GCV-optimal fits, GLM

B.4 Canonical Link

For the canonical link, $w_{Fi} = g(\eta_i) = v(\mu_i)$ and $u_{Fi} = (y_i - \mu_i)/w_{Fi}$. Also, since $d\theta_i/d\beta = x_i$ and $d^2\theta_i/d\beta\beta^t = 0$, it follows that $\mathrm{E}\mathcal{H}(\beta) = \mathcal{H}(\beta)$. So Newton-Raphson and Fisher scoring are equivalent.

B.5 Confidence Intervals

Normal-approximation 100c% confidence intervals on the mean response are given by

$$G\left(x\beta \pm \Phi_{1-(1-c)/2}\sqrt{x^t V x}\right), \tag{B-30}$$

where Φ is a standard normal quantile, V is the estimated parameter variance matrix, and x is a row of an X matrix corresponding to the desired level. For the basis implementation, this is

$$x = \left(f_1(v), \ldots, f_p(v)\right). \tag{B-31}$$

B.6 Bernoulli Response

For example, suppose the response $Y \in \{0,1\}$ is Bernoulli with $\Pr[Y = 1] = \mu = 1 - \Pr[Y = 0]$. The Bernoulli distribution is a member of the exponential family, Eq. B-2, with the particular form

$$f(y) = \mu^y(1-\mu)^{1-y} = \exp\left[y \log\frac{\mu}{1-\mu} + \log(1-\mu)\right]. \tag{B-32}$$

So $a = 1$, $c = 0$, and there is no nuisance parameter. Furthermore, $\theta = \log\left(\mu/(1-\mu)\right)$ and $\mu = 1/(1+e^{-\theta})$, and so $b(\theta) = -\log(1-\mu) = \log(1+e^{\theta})$. Note that $E[Y] = b'(\theta) = e^{\theta}/(1+e^{\theta}) = \mu$ and $\mathrm{Var}[Y] = v(\mu) = b''(\theta) = e^{-\theta}/(1+e^{-\theta})^2 = \mu(1-\mu)$ as expected.

With $\eta = \theta$ and $\mu = G(\theta)$, we see that the canonical link for Bernoulli response is the logistic cumulative distribution function (CDF) $G(\eta) = 1/(1+e^{-\eta}) = \mu$. Note that $g(\eta) = \mu(1-\mu) = v(\mu)$, so $w_{Fi} = \mu_i(1-\mu_i)$ and $u_{Fi} = (y_i - \mu_i)/\left(\mu_i(1-\mu_i)\right)$. The resulting model is logistic regression, or the logit model.

For an arbitrary link CDF G, we take $\eta = X\beta$, $\mu = G(X\beta)$, $v(\mu) = \mu(1-\mu)$, and $g(\eta) = g(X\beta)$.

Use of the standard normal CDF $G = \Phi$ with probability density function (PDF)

$g = \phi$ gives the probit model.

Because the likelihood function is $L = \prod \mu_i^{y_i} \cdot (1 - \mu_i)^{1-y_i}$, we have $L_{\text{full}} = 1$ for the Bernoulli model, and then the deviance is $\Delta = -2 \log L$.

As an example, consider the usual 2-parameter model with predictor $(x_1, \ldots, x_n)$ and response $(y_1, \ldots, y_n)$. The increment $\delta = (d_0, d_1)$ is the solution of $M\delta = A$, where

$$M = X^t W X = \begin{bmatrix} \sum w_i & \sum w_i \, x_i \\ \sum w_i \, x_i & \sum w_i \, x_i^2 \end{bmatrix} \quad \text{and} \quad A = X^t W U = \begin{bmatrix} \sum w_i \, u_i \\ \sum w_i \, u_i \, x_i \end{bmatrix}. \tag{B-33}$$

To do the Fisher update of Section B.2.5, calculate the linear response $\eta_i = b_0 + b_1 x_i$, mean $\mu_i = G(\eta_i)$, derivative $g_i = g(\eta_i)$, variance $v_i = \mu_i(1 - \mu_i)$, transformed response $u_i = u_{Fi} = (y_i - \mu_i)/g_i$, weight $w_i = w_{Fi} = g_i^2/v_i$, and weighted transformed response $w_i u_i = w_{Fi} \, u_{Fi} = (y_i - \mu_i) g_i / v_i$.

For the Newton-Raphson update, $w_{Di} = (y_i - \mu_i)(g_i' v_i - g_i^2 v_i')/v_i^2$ and $w_{Ni} = w_{Fi} - w_{Di}$ and $u_i = u_{Ni} = w_{Fi}/w_{Ni}(y_i - \mu_i)/g_i$. Then $w_i = w_{Ni}$ and $w_i u_i = w_{Ni} \, u_{Ni} = w_{Fi} \, u_{Fi}$.

For the canonical link $w_i = w_{Fi} = g_i = v_i$ and $w_i u_i = w_{Fi} \, u_{Fi} = y_i - \mu_i$.

Intentionally left blank.

Appendix C. B-splines

A degree-d B-spline basis $(B_{1,d}, \ldots, B_{n,d})$ of dimension n uses $n + d + 1$ knots $(t_1, \ldots, t_{n+d+1})$.

There are $k = n - d - 1$ internal knots (internal to the data) and $2d + 2$ boundary knots, $d + 1$ on each side. There are no internal knots if $k = 0$ and $n = d + 1$, and in general $k \geqslant 0$ so $n > d$. The left knots are $(t_1, \ldots, t_{d+1})$, the internal knots, when $n \geqslant d + 2$, are $(t_{d+2}, \ldots, t_n)$, and the right knots are $(t_{n+1}, \ldots, t_{n+d+1})$.

On the data range $[x_0, x_1]$ internal knots are evenly spaced so $t_i = x_0 + \frac{i-d-1}{k+1}(x_1 - x_0)$ for $i = d + 2, \ldots, n$. This also accounts for $t_{d+1} = x_0$ and $t_{n+1} = x_1$.

Compact knots are constructed by replicating knots at the boundary, so $t_1 = \cdots = t_{d+1} = x_0$ and $t_{n+1} = \cdots = t_{n+d+1} = x_1$.

Uniform knots are constructed by repeating the uniform spacing for all knots, so $t_i = x_0 + \frac{i-d-1}{k+1}(x_1 - x_0)$ for $i = 1, \ldots, n + d + 1$.

This is the Cox-de Boor recursion for B-spline basis functions:

$$\begin{aligned}
B_{i,0}(x) &= \begin{cases} 1\,, & t_i \leqslant x < t_{i+1} \\ 0\,, & \text{otherwise} \end{cases}\,, \\
&\quad i = 1, \ldots, n + d \quad ; \\
B_{i,j}(x) &= \frac{x - t_i}{t_{i+j} - t_i} B_{i,j-1}(x) + \frac{t_{i+j+1} - x}{t_{i+j+1} - t_{i+1}} B_{i+1,j-1}(x)\,, \\
&\quad j = 1, \ldots, d \quad , \quad i = 0, \ldots n + d - j\,.
\end{aligned} \tag{C-1}$$

This is the relation for derivatives of B-spline basis functions:

$$\frac{d}{dx} B_{i,j}(x) = \frac{j}{t_{i+j} - t_i} B_{i,j-1}(x) - \frac{j}{t_{i+j+1} - t_{i+1}} B_{i+1,j-1}(x)\,. \tag{C-2}$$

Example graphs follow. All spline bases have dimension $n = 8$. Figures C-1 and C-2 demonstrate uniform and compact knots, respectively, for bases of degree $d = 0, 1, 2$, and 3. Figures C-3 and C-4 demonstrate uniform and compact knots, respectively, for bases of degree $d = 3$ with derivative orders $p = 0, 1, 2$, and 3.

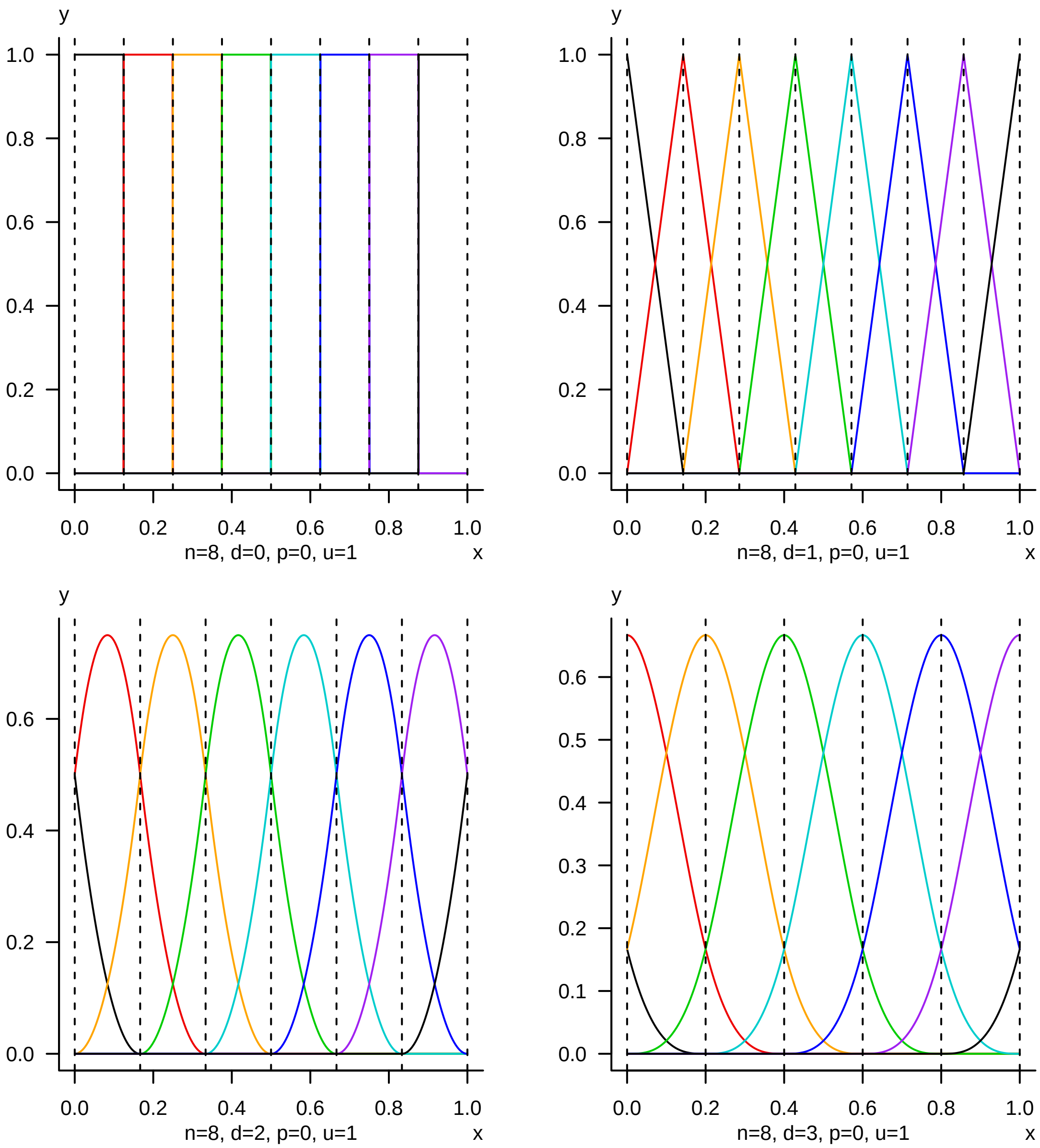

Fig. C-1 B-spline basis, uniform

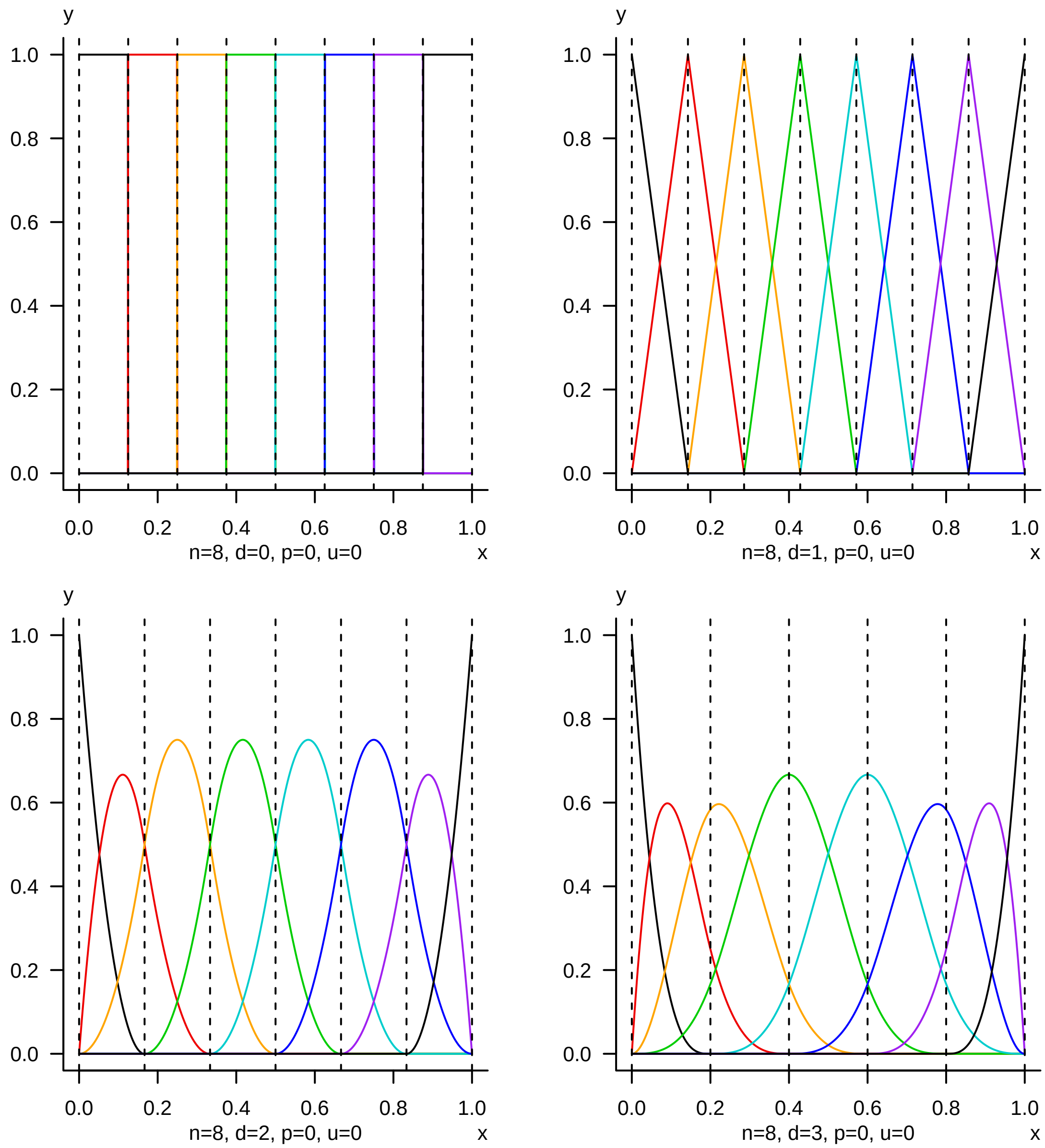

Fig. C-2 B-spline basis, compact

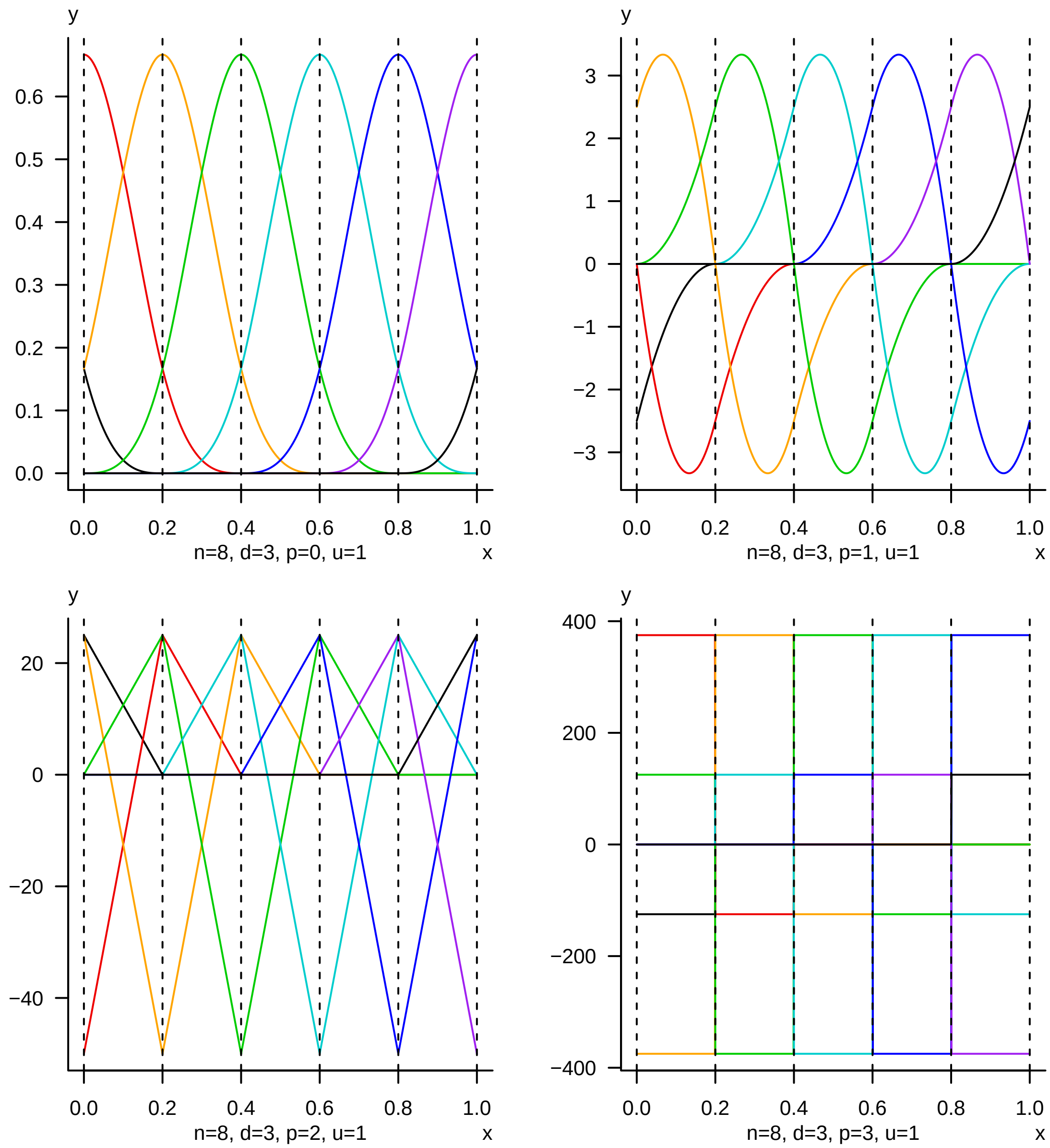

Fig. C-3 B-spline basis derivatives, uniform

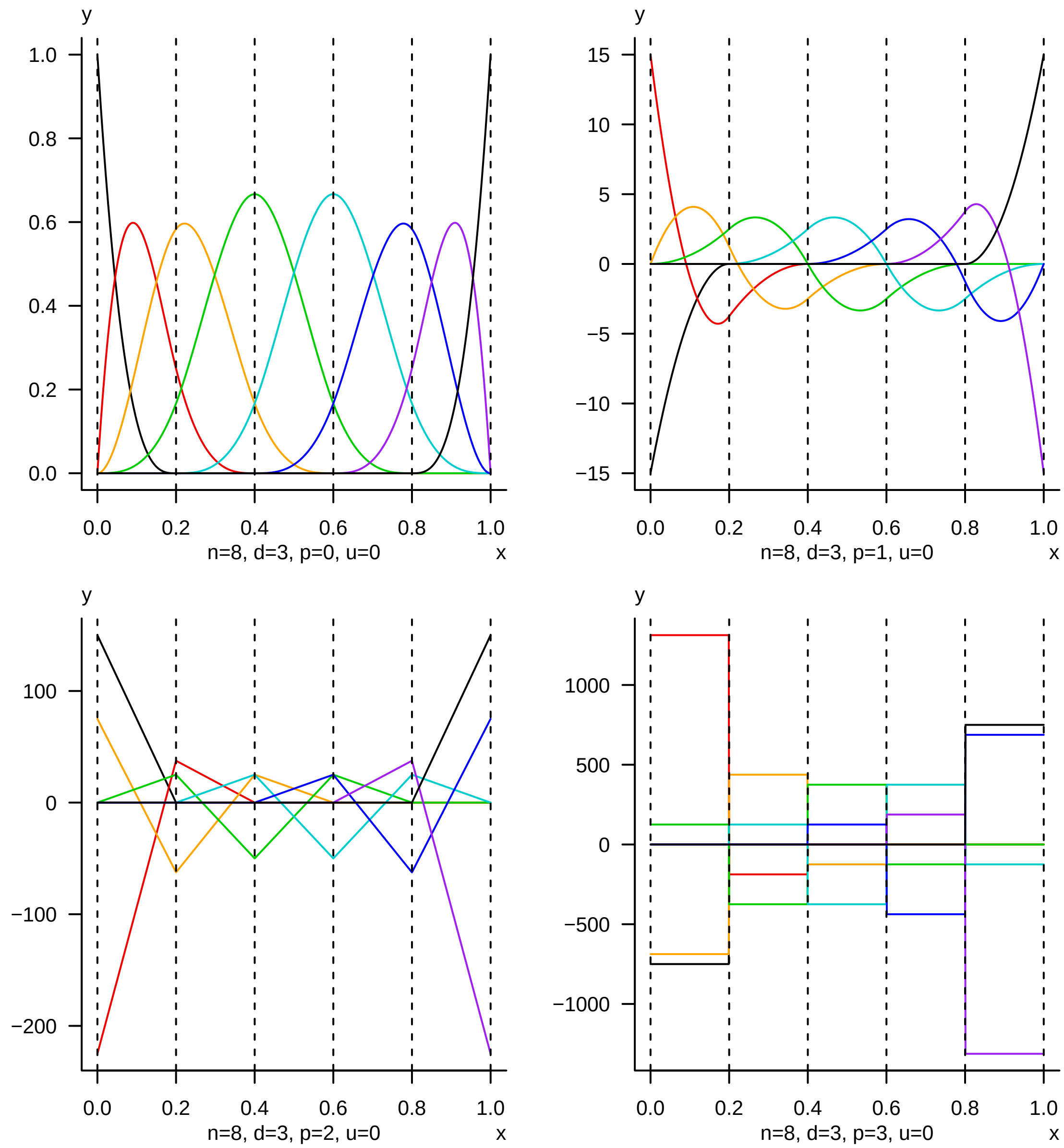

Fig. C-4 B-spline basis derivatives, compact

Appendix D. Matrix Differentiation Operators

Consider that $y = f(x)$, so y is a function of x, and n values of $z_i = (x_i, y_i)$ for $i = 1, \ldots, n$ are given, where each $x_i < x_{i+1}$. Let $f_{k,p}$ be the degree-p interpolating polynomial through the $p+1$ consecutive points $(z_k, \ldots, z_{k+p})$, so that $f_{k,p}(x_i) = y_i$ for $k \leqslant i \leqslant k+p$. Interpolating polynomials satisfy the recursion

$$f_{k,p}(x) = \frac{(x - x_k) f_{k+1,p-1}(x) + (x_{k+p} - x) f_{k,p-1}(x)}{x_{k+p} - x_k} \,. \tag{D-1}$$

Define the Newton basis polynomials $n_{k,j}$ by

$$\begin{aligned} n_{k,0}(x) &= 1\,, \quad 1 \leqslant k \leqslant n \\ n_{k,j}(x) &= \prod_{i=k}^{k+j-1} (x - x_i)\,, \quad 1 \leqslant k \leqslant n\,, 1 \leqslant j \leqslant n-k\,, \end{aligned} \tag{D-2}$$

so $n_{k,j}(x_i) = 0$ for $j \geqslant 1$ and $k \leqslant i \leqslant k+j-1$. Note that $n_{k,j}$ has degree j and the leading coefficient 1, so $n_{k,j}(x) = x^j +$ lower-degree terms. Write the polynomial as

$$f_{k,p}(x) = \sum_{j=0}^{p} a_{k,j} n_{k,j}(x) \tag{D-3}$$

so that

$$\begin{aligned} f_{k,0}(x) &= a_{k,0} n_{k,0}(x)\,, \quad 1 \leqslant k \leqslant n \\ f_{k,j}(x) &= a_{k,j} n_{k,j}(x) + f_{k,j-1}(x)\,, \quad 1 \leqslant j\,, \end{aligned} \tag{D-4}$$

and the leading coefficient of $f_{k,j}$ is $a_{k,j}$. Then, because of Eqs. D-1 and D-4, the coefficients obey

$$\begin{aligned} a_{k,0} &= y_k \\ a_{k,j} &= \frac{a_{k+1,j-1} - a_{k,j-1}}{x_{k+j} - x_k} \,. \end{aligned} \tag{D-5}$$

The coefficient sequence is

$$
\begin{aligned}
&(\, a_{k,0}\, ,\ a_{k,1}\ \ a_{k,2}\, ,\ a_{k,3}\, ,\ a_{k,4}\, ,\ \cdots\,) = \\
&\left(y_k\, ,\ \frac{y_{k+1}-y_k}{x_{k+1}-x_k}\, ,\ \frac{\frac{y_{k+2}-y_{k+1}}{x_{k+2}-x_{k+1}} - \frac{y_{k+1}-y_k}{x_{k+1}-x_k}}{x_{k+2}-x_k}\, ,\ \frac{\frac{\frac{y_{k+3}-y_{k+2}}{x_{k+3}-x_{k+2}} - \frac{y_{k+2}-y_{k+1}}{x_{k+2}-x_{k+1}}}{x_{k+3}-x_{k+1}} - \frac{\frac{y_{k+2}-y_{k+1}}{x_{k+2}-x_{k+1}} - \frac{y_{k+1}-y_k}{x_{k+1}-x_k}}{x_{k+2}-x_k}}{x_{k+3}-x_k}\, , \right. \\
&\left. \frac{\frac{\frac{\frac{y_{k+4}-y_{k+3}}{x_{k+4}-x_{k+3}} - \frac{y_{k+3}-y_{k+2}}{x_{k+3}-x_{k+2}}}{x_{k+4}-x_{k+2}} - \frac{\frac{y_{k+3}-y_{k+2}}{x_{k+3}-x_{k+2}} - \frac{y_{k+2}-y_{k+1}}{x_{k+2}-x_{k+1}}}{x_{k+3}-x_{k+1}}}{x_{k+4}-x_{k+1}} - \frac{\frac{\frac{y_{k+3}-y_{k+2}}{x_{k+3}-x_{k+2}} - \frac{y_{k+2}-y_{k+1}}{x_{k+2}-x_{k+1}}}{x_{k+3}-x_{k+1}} - \frac{\frac{y_{k+2}-y_{k+1}}{x_{k+2}-x_{k+1}} - \frac{y_{k+1}-y_k}{x_{k+1}-x_k}}{x_{k+2}-x_k}}{x_{k+3}-x_k}}{x_{k+4}-x_k}\, ,\ \cdots \right) .
\end{aligned}
\tag{D-6}
$$

For evenly spaced x_j with $x_{j+1} - x_j = h$ and $x_{k+j} - x_k = jh$, we have

$$
\begin{aligned}
a_{k,0} &= y_k \\
a_{k,j} &= \frac{a_{k+1,j-1} - a_{k,j-1}}{jh}\, ,
\end{aligned}
\tag{D-7}
$$

so

$$
\begin{aligned}
&(\, a_{k,0}\, ,\ a_{k,1}\ \ a_{k,2}\, ,\ a_{k,3}\, ,\ a_{k,4}\, ,\ \cdots\,) = \\
&\left(y_k\, ,\ \frac{y_{k+1}-y_k}{h}\, ,\ \frac{y_{k+2}-2y_{k+1}+y_k}{2h^2}\, ,\ \frac{y_{k+3}-3y_{k+2}+3y_{k+1}-y_k}{6h^3}\, , \right. \\
&\left. \frac{y_{k+4}-4y_{k+3}+6y_{k+2}-4y_{k+1}+y_k}{24h^4}\, ,\ \cdots \right) .
\end{aligned}
\tag{D-8}
$$

For evenly spaced x, the coefficients are

$$
a_{k,j} = \frac{1}{j!h^j} \sum_{i=0}^{j} (-1)^{j+i} \binom{j}{i} y_{k+i}\, . \tag{D-9}
$$

In general, the numerical derivative of order p coincides with the p^{th} derivative $f_{k,p}^{(p)}$ of $f_{k,p}$,

$$
f_{k,p}^{(p)}(x) = p!\, a_{k,p}. \tag{D-10}
$$

There are $n - p$ sequences $(z_k, \ldots, z_{k+p})$ and therefore $n - p$ values of $f_{k,p}^{(p)}$ for $1 \leqslant k \leqslant n - p$.

Now, express the differentiation operator as a matrix.

Define $\nu(j)$, the first-order difference matrix of dimension $k \geqslant 2$, size $(j-1) \times j$, by $\nu(j)_{i,i} = -1$ and $\nu(j)_{i,i+1} = 1$ for $1 \leqslant i \leqslant j-1$.

$$\nu(j) = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}. \tag{D-11}$$

Then

$$\nu(j) \begin{bmatrix} u_1 \\ \vdots \\ u_j \end{bmatrix} = \begin{bmatrix} u_2 - u_1 \\ \vdots \\ u_j - u_{j-1} \end{bmatrix}, \tag{D-12}$$

and ν serves to evaluate the numerator of Eq. D-5.

Define $\delta(u, j)$, the lag-j (square) diagonal difference matrix of dimension $n - j$ for $u = (u_1, \ldots, u_n)$ and $1 \leqslant j \leqslant n-1$ by $\delta(u, j)_{i,i} = u_{j+i} - u_i$ for $1 \leqslant i \leqslant n-j$

$$\delta(u, j) = \begin{bmatrix} u_{j+1} - u_1 & & \\ & \ddots & \\ & & u_n - u_{n-j} \end{bmatrix}. \tag{D-13}$$

So δ serves to evaluate the denominator of Eq. D-5.

Then matrix derivative operators for the n-vector x are given recursively by

$$\begin{aligned} D_1 &= \delta(x, 1)^{-1}\, \nu(n) \\ D_j &= j\, \delta(x, j)^{-1}\, \nu(n-j+1)\, D_{j-1}, \end{aligned} \tag{D-14}$$

where the factor of j serves to evaluate the factorial in Eq. D-10.

The matrix D_P evaluates the p^{th} derivative.

Based on points $(z_k, \ldots, z_{k+p})$ for $1 \leqslant k \leqslant n-p$, the individual derivative value are

$$f^{(p)}_{k,p}(x) = (D_p\, y)_k\,. \tag{D-15}$$

The derivative vector is $D_p\, y$.

$$D_p y = \begin{bmatrix} f_{1,p}^{(p)}(x) \\ f_{2,p}^{(p)}(x) \\ \vdots \\ f_{n-k,p}^{(p)}(x) \end{bmatrix}, \tag{D-16}$$

and its norm is approximately

$$\int \left(f^{(p)}(x)\right)^2 dx \simeq \|D_p y\|^2 = y^t D_p^t D_p y\,. \tag{D-17}$$

R language code for evaluating D_p is easy.

```
D <- diff(diag(n)) / diff(x, lag=1)
if (p>1) for (j in 2:p)
    D <- j * (diff(diag(n - j + 1)) / diff(x, lag=j)) %*% D
```

Here, `diff(diag(n))` is $\nu(n)$ and `diff(x, lag=j))` is $\delta(x, j)$.

List of Symbols, Abbreviations, and Acronyms

AIC	Akaike information criterion
CB	Chang-Bodt
CDF	cumulative distribution function
GCV	generalized cross-validation
GLM	Generalized Linear Model
IID	independent and identically distributed
IRLS	iteratively reweighted least squares
OCV	ordinary cross-validation
PDF	probability density function
QR	quantal response
SSE	sum of squared errors

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
2 (PDF)	DIRECTOR US ARMY RESEARCH LAB RDRL CIO L IMAL HRA MAIL & RECORDS MGMT
1 (PDF)	GOVT PRINTG OFC A MALHOTRA
1 (PDF)	DIR US ARMY EVALUATION CTR HQ TEAE SV P A THOMPSON
1 (PDF)	DIR USARL RDRL SL P BAKER
20 (PDF)	DIR USARL RDRL DPW T HOLDREN RDRL SL D BAYLOR P DISALVO T STADTERMAN RDRL SLB R BOWEN G MANNIX RDRL SLB D J COLLINS J EDWARDS R GROTE L MOSS E SNYDER RDRL SLB E M MAHAFFEY RDRL SLB G N ELDREDGE RDRL SLB S R DIBELKA C KENNEDY M PERRY R SAUCIER G SAUERBORN RDRL SLB W S SNEAD RDRL SLE R FLORES

Intentionally left blank.